AMERICAN UNIVERSITY OF BEIRUT

INVENTORY SYSTEM WITH SCHEDULED DEMAND AND DISTRIBUTED SUPPLY

by

SASHA JOHNNY GHAYA

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Graduate Program in Computational Science
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
May 2024

AMERICAN UNIVERSITY OF BEIRUT

INVENTORY SYSTEM WITH SCHEDULED DEMAND AND DISTRIBUTED SUPPLY

by

SASHA JOHNNY GHAYA

Approved by:

Dr. Victor Araman, Professor Suliman S. Olayan School of Business	Advisor
Dr. Maher Nouiehed, Assistant Professor Industrial Engineering and Management	Member of Committee
Dr. Siamak Taati, Assistant Professor Mathematics	Member of Committee

Date of thesis defense: May 7, 2024

Acknowledgements

As I write this acknowledgment, I realise that I have so many things to be thankful for and that it is not easy to put so many emotions into words. It goes as follows:

My mom Rola, my walking translation engine, my go-to person when I need a confidence boost, a hug, a laugh, or a shoulder to cry on (mind you this was the story of my life these last few weeks). My dad Johnny, my private chef, my bug killer, and my mentor who taught me everything I know about sports and soccer (yes believe it or not, I can spot an offside before a referee or a VAR does!) You both made this journey bearable and possible when I felt I couldn't go on anymore. I'm forever grateful for the laughter, the inside jokes and for having you in my life through thick and thin.

My grandmother Jamal who is my favorite chef and year-around Santa, my grandmother Raja who is my roommate and my soft spot. My godparents, auntie Aline from whom I inherited my crazy curls and her husband Elie who is a second father to me. My aunt Sisi who pampers me. And my grandfather Emile, my guardian angel whom I miss dearly and who always made me feel special. I'm grateful for your loving presence.

My lifelong neighbors Hana and Noha who always spoil me and support me even when I'm wrong. I'm so lucky to have you in my life.

My neighbor/friend Camille Junior who is literally the brother I never had, my personal driver and the one I enjoy having food with. Thank you for bearing with me when I am down, for your selfless kindness and for sharing my passion for anime.

My advisor Dr. Araman, from whom I learned a lot and I mean a lot not only academically but on the personal level. Thank you for your dedication, your sense of humour and most importantly your kindness. I hope I can someday be a similar version of you.

My committee members, Dr. Nouiehed and Dr. Taati, thank you for your valuable remarks, contribution and guidance.

My friends, my partners in crime, Ali and Melissa. I will never forget the laughters, the long chats and especially the frequent meltdowns we had recently. I wish you the best life can give.

Finally, I would like to thank Father Georges for his support and spiritual guidance and most importantly I thank God for this achievement and for all the things to come. I know He was, is and will be there for me every step of the way. One word: Blessed.

Abstract
of the Thesis of

Sasha Johnny Ghaya for Master of Science
Major: Computational Science

Title: Inventory System with Scheduled Demand and Distributed Supply

We consider a simple supply chain model constituted of a retailer facing end-consumers' demand while being supplied by a large number of suppliers. We model the retailer as a single server queue while the independent suppliers are assumed to form an infinite server system. Many applications can fit this setting. We focus in this work on the case where the market demand is predictable (i.e., orders arriving following a deterministic sequence) while concentrating the uncertainty on the supply side through the processing time of each "server". In this setting, suppliers decide first on their capacity level followed by the retailer who decides on the adequate base-stock level. From a queueing perspective, suppliers can be represented by a $D/G/\infty$ queue. The retailer's queue turns out to be an $S/D/1$ queue (the S denotes a scheduled traffic as defined in Araman et al. (2021)), where the positive perturbation is the supplier's processing time. To analyze this system, we consider first the centralized system as a benchmark where the retailer sets both the capacity level as well as the base-stock level. For the decentralized setting, we consider two cases. In the first one, all suppliers are under one supply function and decide their capacity levels as one entity. In the more interesting case, we assume that suppliers decide individually on their capacity. The objective function is the inventory cost rate that suppliers and retailer are each minimizing. Even under exponential perturbations, the problem is intractable. We therefore suggest, through an asymptotic analysis, a full characterization of the optimal centralized and Nash solutions under a heavy traffic regime. Moreover, we perform a numerical analysis to validate these approximations through a Monte Carlo simulation.

Table of Contents

Illustrations

Tables

Chapter 1

Introduction

The gig economy has emerged as a prominent force in today's business landscape, significantly impacting the way companies operate and the nature of employment. By definition, the **Gig economy** is a labor market that relies heavily on freelancers and independent workers to perform temporary tasks/ jobs and provide short-term services rather than traditional full-time employment. Typically, workers and customers are linked through an online or digital platform that would match the worker to a task or a customer. These platforms employ millions of workers every year, and it is expected that the economy reaches $873 billion by 2031. The gig economy offers the workers flexibility in their schedules and allows them to select tasks that match their interests and abilities. Moreover, it provides them with a source of income often to complement their primary employment earnings. It encourages autonomy seeking and offers a diverse work experience. On the other hand, it provides customers with more customized services and innovative solutions, often at competitive prices. Additionally, platform-based services gives customers easier access to services that meet their demand.

Often such mode of employment is prevalent within the service sector, notably in transportation. Uber and Lyft are prime examples of how businesses have leveraged the gig economy in the context of hail-riding. These platforms connect users with independent drivers who can provide on-demand transportation services and thus offer several advantages, including ease of booking, cashless transactions, and reliable service. Clearly, such crowdsourcing-like activities have extended to diverse other fields including healthcare, design, professional services, and beyond.

In recent times, flexible employment strategies have also been implemented within the manufacturing sector. Some companies, like SOKO and Bokksu, are now establishing what is known as a distributed supply chain designed around numerous small local suppliers (often specialized artisans) instead of relying on large global manufacturing plants. SOKO, a Nairobi-based company connects businesses directly with independent artisans and suppliers in emerging markets, allowing for transparent and ethical sourcing of unique handmade products. These work opportunities not only empower local artisans but also provide businesses such as SOKO with a diverse range of high-quality products produced locally. On the other hand, Bokksu specializes in the distribution of Japanese snacks and treats. The platform partners with local snack makers in Japan and then delivers these curated snacks directly to customers worldwide. This direct connection between independent snack makers and consumers not only ensures the authenticity and quality of the products but also promotes cultural exchange and appreciation. Both Soko and Bokksu exemplify how the gig economy has facilitated direct connections between businesses and independent suppliers, offering a range of benefits such as streamlined procurement processes, access to unique goods, and the opportunity to support independent artisans and local businesses.

In view of the above, our aim in this project is to analyze simple supply chains that rely on distributed systems, understand the complexity induced by the large number of players involved and finally contribute to how to make such systems more efficient. For that, we consider a two-tier supply chain constituted of a make-to-stock retailer and a distributed system of suppliers that we denote by the supply function. The retailer faces a stream of orders from customers that need to be met from its current inventory (stock). The retailer relies on the supply function to replenish its inventory when needed. The latter through the large number of independent and active suppliers offers a major flexibility to the retailer but induces also a great uncertainty in the system. In our aim to understand these conflicting features and their impact on the supply chain, we restrict the uncertainty in the system to the supply side, specifically through the processing times of orders and focus on settings where the demand stream is (quasi) deterministic. This latter assumption is in line with a number of current applications specifically in the context of Gig economy. Indeed, recent years have seen a rise in companies with subscription models or appointment-based systems. This type of models gives the company a higher visibility on its upstream demand and allows them to reduce the unpredictability of their orders' arrivals.

Our objective is to understand the intricate relationship between a central retailer and a distributed supply system. In this context, the research questions we ask revolve around the operational and tactical considerations of both the retailer and suppliers as independent cost minimizing entities. In particular, the retailer is deciding on the replenishment quantity which is reduced to setting a base-stock level. As for suppliers they face tactical decisions regarding the capacity level they will be allocating to the retailer.

We adopt a queueing theoretic approach to model a competitive supply chain in the presence of large number of suppliers. We rely on asymptotic approximations governed by a heavy traffic regime that seems natural in the context of the gig economy. Our main contributions are manifold, of which the most important are: *i*) a tailored queuing theoretic model for a supply chain with distributed system that is amenable for (asymptotic) analysis in a game-theoretic context. Moreover, *ii*.) the handling of the game theoretic dynamics using mean field games in the context of competitive supply chains that involves a large number of players. Such approach would allow firms to simplify very complex dynamics and obtain a full characterization of the equilibrium, leading eventually to the design of mechanisms (contracts) that allow to reach the first best solution. In addition, *iii*.) it is worth noting that even the centralized case where one player is making all the decisions, the problem is quite complex given the infinite dimensional optimization problem involved. We handle this issue by showing that the infinite number of decisions can be represented at optimality by an empirical distribution which *always* be mapped to a uniform distribution. Finally, *iv.*) many managerial insights can be obtained from our analysis regarding a relevant and contemporary filed that has been overlooked in the literature. Before introducing the model in Chapter 3, Chapter 2 is devoted to the literature review that positions the paper in the context of a large stream of papers on competitive supply chains but that lacks exhaustive coverage of distributed systems. Our main results are obtained in Chapter 4. We start by analyzing a centralized chain looking for the first-best solution. In the decentralized setting we consider two cases. First, the case where the entire supply function is centralized represented by one large entity. More importantly, we analyze next the case of a fully decentralized system where each supplier is an independent entity. In these two cases, we look for the Nash equilibrium and measure the inefficiency relatively to the first-best solution. For the fully decentralized case, we reduce the dynamics complexity through the use of *mean field equilibrium*. We show the existence of such equilibrium and obtain a full characterization of the solution. In Chapter 5 we suggest a mechanism design through a class of contracts that if offered by the retailer, would lead to a full coordination and efficiency of the supply chain. These contracts not only achieve a first best solution, they do so by improving all the agents performance. Finally, in Chapter 6 we consider Stackleberg's games, where one of the agents is a leader. This gives the leader an advantage as they optimize their decision variables given the response function of the remaining agents. Finally, in Chapter 7, we perform a numerical analysis through Monte Carlo simulation to verify our results. We should specify that the theoretical results obtained throughout are coupled with a numerical analysis that allows us to either extract the relevant managerial insights, verify optimal solutions or calculate numerically non-tractable solutions.

Chapter 2

Literature Review

Over the last few decades, supply chain management has been the center of extensive research which came to complement and build upon a vast literature on logistics and in particular on inventory management that date back to a paper published in 1913 by Ford Whitman Harris as reported by [1], followed by many influential work such as the seminal paper of [2]. This literature views the problem from the perspective of one firm, and focuses typically on obtaining optimal policies regarding inventory, transportation, capacity, among many other logistical drivers. We refer to two recent reviews by [3] and [4] on Capacity and Inventory Management for the first and on different types of Capacity Management for the second. We also mention [5] who analyze a setting of coordinated inventory and capacity management through process simulation.

Supply chain management on the other hand is primarily interested in the interaction between multiple players who either belong to different supply chain stages (supplier and retailer) or/and to the same stage (supplier and multiple retailers). In this work we are interested in the concept of distributed supply where a retailer relies on a very large number of small suppliers to produce its products and meet market demand.

The literature on supply chain management has been tackling a number of problems stemmed from these multi-agents interactions and the inefficiencies they generate caused by a variety of issues such as (the lack of) information sharing (e.g., Bullwhip effect ([6]) or double marginalization effects and contract design (see, [7]). We contribute to this literature by introducing a model for a distributed supply system and aim at understanding its implications on the supply chain, in particular on the retailer's performance and on customers' satisfaction. We are specifically interested in analyzing the connection between a decentralized capacity decision from the supply side and an inventory decision from the retail side.

A great majority of the supply chain management literature considers discrete time settings such as one or two periods problems. We refer again to [7] for a comprehensive review and mention couple of other papers that are connected to our work and representatives of this literature. The work of [8] introduces a two-stage supply chain where a supplier and a retailer choose each their base-stock policies independently in the context of two different information tracking settings.

A Nash equilibrium is obtained for each setting and the solutions are compared to the centralized case. The results confirm that competition increases operational inefficiency. Other papers like [9] and [10] model a more complex system involving a single supplier and many retailers. Both papers prove the existence of a Nash equilibrium that is shown to be different than the first best solution. We also mention [11] who consider a supply chain constituted of a number of manufacturers competing for one retailer's business. Manufacturers start each by offering a contract to the retailer and accordingly the retailer responds by setting the market prices which would determine the demand function. Three different types of contracts are considered which lead to different dynamics from the case of a simple one-manufacturer and one-retailer supply chain.

There is a stream of supply chain papers that emphasize the time dynamics through the use of continuous time models, often relying on queueing systems to represent the various players involved. An early example is [12], who introduce a customers' balking model in the context of a toll company seeking to design regulations that would allow it to maximize profit and improve service. A more recent work is [13] who suggest an incentive-compatible priority pricing scheme that maximizes social welfare. The provider decides on the price followed by wait-sensitive customers who decide to join the system or not and at what priority level. Among those using queueing-based modeling, the closest paper to ours is [14] that models a make-to-sock retailer facing a Poisson process demand function. Like us, the retailer applies a base-stock policy and continuously replenishes its demand from a single supplier however the latter is modeled as a single server queue. The paper obtains a Nash equilibrium in the decentralised setting that is compared to the solution of the centralised problem. The paper suggests a family of contracts based on transfer payments that coordinates the system.

There are many differences between the current work and [14]. Structurally speaking, we consider a distributed supply rather than a single supplier and account for the processing time that occurs at the retailer level through an additional (single server) queue. As opposed to the typical Poisson process demand used in such models, we consider a deterministic demand in order to further highlight the suppliers' uncertainty. This assumption requires the use of a new set of queueing results away from those of a standard $M/M/1$ model that is typically used in the literature (e.g., [14]). The deterministic demand assumption can be suitable in many real-world applications when a renewal or a Poisson process assumption might not at all be adequate. For instance, many systems are increasingly requiring a pre-defined schedule to receive orders (e.g., Amazon fulfillment center) and others rely highly on appointments. As a result, the entire analysis in this work is different than that of [14].

Another stream of recent literature that is relevant to ours is that concerned with the operations of Gig economy and crowdsourcing. An illustration of it is [15], where an on-demand platform is studied. In their platform, customers exhibit sensitivity to delays and 'independent' agents decide whether or not to participate in the platform. Taylor's analysis assumes that job allocation mechanism is a given factor and primarily delves into the repercussions of customer delay sensitivity and worker independence on optimal prices and wages. Another relevant work is by [16], where they model an online labor platform's operations. The demand are the customers that submit jobs to be fulfilled in a certain amount of time while the supply is the workers in need of a job. They are connected through a platform that assigns jobs to workers and charges them a certain price. They propose a pricing and allocation policies that would maximize revenues and minimize unpredictability in workers' profits. They show the optimality of their policy through an asymptotic analysis and examine its performance through a discrete event simulation. Similar to ours, both these works model there platform through queueing models, but they are both concerned about pricing and job allocation.

Another relevant stream of papers is the one concerned with stochastic games involving a large number of interacting players. Some of them explore the "Mean-Field equilibrium", a distinctive form of Nash equilibrium where each player's strategy is influenced by the population's behavior. [17] and [18] draw inspiration from statistical physics and introduce Mean Field Games (MFG), where the key concept is to utilize the mean field distribution associated with the limiting scenario of infinite players. This simplifies the model and its analysis, as it replaces the interaction between all individuals with the interaction of one player and the mean field distribution. One application of this equilibrium is [19] where they survey the literature and combine MFG with reinforcement learning. Another application is [20], where the authors describe a system of repetitive ad-exchange auctions.The publisher maximize her pay-off and advertisers optimizes their bidding strategy. Their main contribution is to provide a new notion (Fluid Mean Field Equilibrium) that combines the standard Mean Field Equilibrium with Stochastic Fluid approximations. Some common factors between our and [20]'s work is the use of infinite queue with general service time. The players (advertisers) are also involved in what we call a vertical game as they compete to optimize their policies and thus increase their profit through an auction. In contrast, our work contributes in making use of the Mean Field equilibrium to allow suppliers competing to optimize their capacity and thus minimize their costs.

Chapter 3

The Model

3.1 Model set-up and Ingredients

Consider a distributed system made of a large number of suppliers working independently of each other. *This supply function* is connected through a mobile technology to a retailer that typically owns the intellectual property (e.g., a jewellery designer). Every season, the retailer introduces new designs and train the suppliers - local artisans - on each one of them before the start of the season. Once the season starts, the retailer receives orders continuously from direct customers and fulfill them from its inventory. The retailer replenishes its inventory by transmitting orders in a distributed manner to the suppliers through mobile technology. These artisans/suppliers produce and deliver back the orders to the retailer who is usually responsible for testing the units before making them available for sales. We disregard in this model any defect rate. When backorders occur at times where there is no current inventory to meet demand, all the players of the supply chain share the costs. We denote by b the unit backorder cost and by $\alpha \in [0, 1]$ the share of the retailer from the backorder cost (with $1 - \alpha$ being the share of the supply function). The retailer also incurs a holding cost h per unit of inventory that is held. As for the suppliers, each one of them needs to decide the capacity to allocate to the retailer and for that they incur a capacity cost which often reflects the missed opportunity of not allocating this capacity for other endeavors available.

The supply function is made of a large number of suppliers/artisans that are independent who process and deliver orders as received from the retailer at their own pace governed by the capacity allocated. It is therefore natural to view this supply function as a multi-server queueing system. In that regard, we specifically consider an infinite server queue given the very large number of artisans typically involved in the distributed systems we are interested in. The queueing system faces an input traffic corresponding to customers demand while the output process represent finished goods meant to replenish the retailer's inventory.

We assume in this work that end customers' demand is constant and deterministic with a rate of arrivals denoted by λ. The demand is met from the retailer's inventory otherwise backlogged. The retailer follows a base-stock policy which is parameterized by S, and replenish orders by relying on the network of suppliers.

Each order received downstream is then automatically transmitted to an idle supplier. Once processed, the order is submitted to the retailer who checks the quality and makes it available for sales. We could view the retailer as a single server queue with input traffic the departure process from the supply function. We will argue later that under some assumptions this queue has no impact on the analysis and one can drop it. The retailer can therefore be reduced to the platform that meets the demand from its inventory while simultaneously ordering a replenishment for these demanded units. The latter policy is a practical execution of an order-up-to policy. We assume that supplier k has full control on the capacity μ_k to allocate to the order received form the retailer. We denote by ξ_n^k the processing time of order n if allocated to supplier k with $\mu_k = 1/\mathbb{E}\xi_1^k$ and assume that for any k, $(\xi_n^k : n \geq 0, k \geq 0)$ is an independent sequence of random variables (rv's) with known distribution governed by μ_k. For most of our analysis we assume that ξ_1^k is exponentially distributed with rate μ_k. In terms of notations, we drop the k from ξ_1^k for a generic supplier. We also denote by $\mathbb{E}_F$ the operator expected value with respect to a distribution F. We drop usually F from the notation when it is clear with respect to which distribution the operator is applied to.

As mentioned above, the problem we are facing is one of large number of artisans and a retailer (supply chain agents) where each one of them independently makes an operational decision. The retailer needs to decide on the base stock policy S, while each supplier k decides on the capacity μ_k to allocate to the retailer. The resulting performance of each agent obviously depends on its decision and all the other agents' decisions. We suggest to structure and analyze this game theoretic setting following "**Mean Field Games**(MFG)". These types of games offer a mathematical framework that lends itself for strategic interactions among a large number of agents. We refer the reader to [19] for a survey on the different types of MGFs with specific example for each type. In this work, we rely specifically on "**static MFGs**" where agents, as opposed to dynamic games, take a single decision with no impact of time. The agents will be looking for a "**Mean Field Nash Equilibrium** (MFNE)" where the equilibrium represents a *stable* state, in the sense that no agent has an incentive to unilaterally deviate from its position when all other agents are assumed to keep theirs. In our model, the final values of μ_k's define a distribution that we denote by Γ which represents the population behaviour. The induced r.v. is denoted by $\hat{\mu}$. As [19] describes, the MFNE is a pair of elements: the individual behaviour and the population behaviour. In our case, these are μ_k^* and Γ^*.

Note that it is extremely hard for the retailer to track each supplier's capacity given their number and the complex setting they operate in. Moreover, the capacity of a supplier might change between one order and the other even though in equilibrium the overall distribution remains the same. Finally, in many cases, the platform announces the order and the order is allocated to the artisan who accepts it first and hence the retailer cannot predict which supplier will do. For all these different reasons, we assume throughout this analysis that the retailer doesn't know the supplier's capacity when allocating an order, and hence from her perspective the capacity allocated to an order is a random variable $\hat{\mu}$ (with μ_k being a realisation of $\hat{\mu}$ for supplier k) which again its distribution is given by the equilibrium distribution, Γ that we denote *the capacity distribution for the supply function*. We also define the quantity

$$\zeta_1 = \mathbb{E}_\Gamma[\xi_1|\hat{\mu}],$$

which can be viewed from the retailer's perspective as the random processing time. The sequence $(\zeta_j : j \geq 1)$ is also i.i.d. Note that even though the retailer doesn't know the specific capacity of each supplier at the time of allocation, she knows the distribution Γ.

The state process depicting the dynamics of the system is the inventory level held by the retailer at any time t. Given the order-up-to policy followed by the retailer, the inventory level is equal to the base stock level S minus the number in system at the supply function, $\mathrm{N_s}(t)$ and minus the number in system at the retailer's single server queue, $\mathrm{N_r}(t)$. Given Γ the capacity distribution for the supply function, the number of units $\mathrm{N_s}(t)$ that are currently being produced by the supply function at time t, can be viewed as the number in system of a $D/G/\infty$ with processing times, ζ and interarrival times, $1/\lambda$. It is shown in [21] that such queueing system admits a steady state distribution. The next lemma characterizes this distribution.

Lemma 1

$$N_s \overset{d}{=} \sum_{j=0}^{\infty} I\left(\lambda\xi_j + U > j\right),$$

where U is a uniform random variable.

This uniform random variable is introduced only for technical reasons in order to guarantee the existence of a steady state. It will not play any role in our analysis. We note from [21] that the quantity $\mathrm{N_s}$ is well defined for a large class of distributions and for instance admits a finite first moment as long as ξ_1 does. For more details on the variable U or in general on this quantity, we refer to [21].

As for the number of units at the retailer at time t, $\mathrm{N_r}(t)$ represents the number in system of an $S/D/1$ queue which is a single server queue where the processing times are deterministic with service rate τ, and the arrival traffic is following a so-called scheduled process. The latter is the traffic obtained from an initially deterministic and constantly paced process that is perturbed at each point. Such process is exactly the departure process from an infinite server queue with deterministic input like the one associated with the large scale distributed system we are considering here. Interestingly, for such $S/D/1$ queue with constant processing times we have that

$$\mathrm{N_r} = \lceil W_r^\rho(t) \rceil,$$

where $W_r^\rho(t)$ is the workload which is the total work accumulated at time t that hasn't been processed yet, and ρ is the utilization of this queue. We refer to [21] for more details on this type of traffic and corresponding queue.

3.2 Supply Chain Costs

All the players in the supply chain be it the suppliers or the retailer are interested in minimizing their costs. We denote by c_k the cost of maintaining a unit capacity for

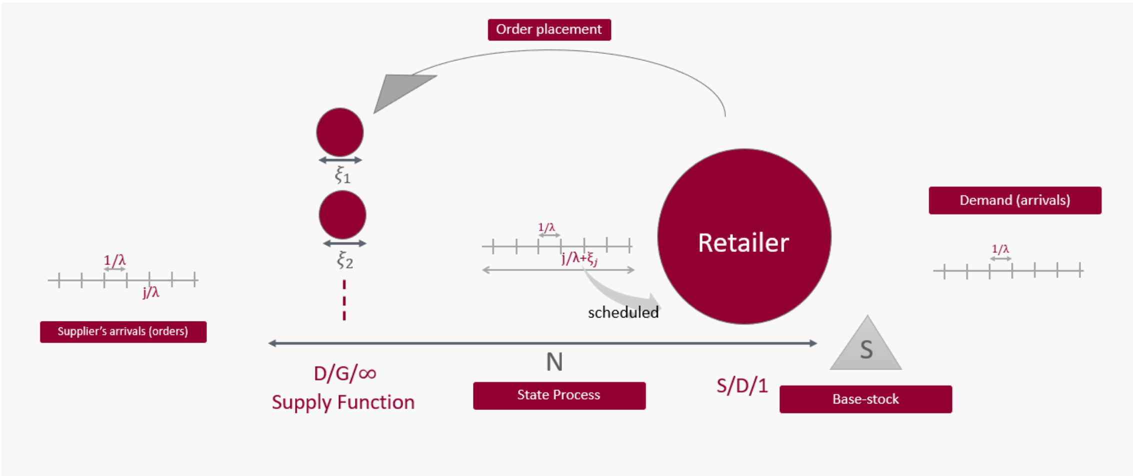


Figure 3.1: Queueing Model

supplier k. Each supplier knows its costs which can be viewed as an opportunity cost. Each artisan's conditions are different and the opportunities available are different and the supplier is giving up on these when offered to work on an order from the retailer. From the retailer's perspective the supplier's cost is a random variable $\hat{c}$ with some given distribution. Set $\boldsymbol{\mu} = (\mu_k : k \geq 1) \in \mathbb{R}_+^\infty$ and let $N \triangleq \mathrm{N_r} + \mathrm{N_s}$. The retailer incurs a holding cost on each unit in inventory and a backlog cost when demand is not directly met. The backlog cost is shared between the retailer and the supply function with α being the retailer's share of the cost. The total inventory cost rate is then given as follows:

$$\mathrm{TC_r}(S;\boldsymbol{\mu}) = h\,\mathbb{E}[S-N]^+ + \alpha\, b\,\mathbb{E}[N-S]^+ \tag{3.1}$$

where h and b are respectively the marginal holding and backlog costs.
Accordingly, the retailer is then solving the following problem:

$$\min_S \mathrm{TC_r}(S;\boldsymbol{\mu}).$$

We move to discuss the supply function problem. Given a supply function capacity $\boldsymbol{\mu}$, the total cost rate of the supply function, $\mathrm{TC_s}(\boldsymbol{\mu};S)$ is given by:

$$\mathrm{TC_s}(\boldsymbol{\mu};S) = (1-\alpha)\, b\,\mathbb{E}[N-S]^+ + \mathbb{E}[\hat{c}\hat{\mu}]. \tag{3.2}$$

In order to argue for the above cost, we start with an observation. Note that the infinite server setting can be viewed as a system that guarantees to the retailer at any time t, and without any delay, that a server is available to take an order. Having that property in mind, we define the index $k(t)$ of the *idle* server/artisan to whom an order will be allocated to would an order be received at time t. WLOG, we assume that $k(t)$ remains constant between orders and changes value only once an order is received. If the order is received at time t then it is allocated to $k(t_-)$ and $k(t)$ takes a new value which is the server's index that the next order will be allocated to.

We can now write the long time average cost incurred by the supply function as follows:

$$\Psi_T = \frac{1}{T}\int_0^T \left((1-\alpha) b \frac{\mu_{k(t)}^{-1}}{\mathbb{E}_\Gamma\, \hat{\mu}^{-1}} \,[N(t) - S]^+ + c_{k(t)}\mu_{k(t)} \right) dt.$$

The first term in the integral is the supply function's share of the backlog and the second term is the cost incurred by the supply function at time t to insure that a capacity μ_k is available to the retailer. Regarding the first term, the factor $\frac{\mu_k^{-1}}{\mathbb{E}_\Gamma\, \hat{\mu}^{-1}}$ represents a way to reward suppliers that select a larger capacity than the average by incurring a smaller share of the backlog cost. Therefore, $\mathrm{TC}_{\mathrm{s}}(\boldsymbol{\mu}; S)$ is obtained as the limit of Ψ_T as $T \to \infty$.

Given an order-up-to policy S followed by the retailer, the previous long time average cost formulation allows us also to identify the optimization problem that each supplier is then solving which is:

$$\min_{\mu_k} \mathrm{TC}_{\mathrm{s}}^k(\boldsymbol{\mu}; S),$$

where

$$\mathrm{TC}_{\mathrm{s}}^k(\boldsymbol{\mu}; S) = (1-\alpha)\, b \frac{\mu_k^{-1}}{\mathbb{E}_\Gamma\, \hat{\mu}^{-1}}\, \mathbb{E}[N - S]^+ + c_k\mu_k. \tag{3.3}$$

The expected value is with respect to all the uncertainties, i.e., the process, the capacity level and the marginal capacity cost.

We could have considered the more general situation where supplier k sets a randomized policy by deciding on a distribution π_k from which a value μ_k is drawn each time supplier k receives an order. However, the fixed value of μ_k is without loss of generality.

3.3 Asymptotic Formulation: Set up

The supply chain optimization problem formulated in the previous section are intractable. It involves infinite number of simultaneous non linear optimizations driven by the heterogeneity of the suppliers reflected by their capacity cost c_k which will drive them to possibly select different capacity levels. Moreover, the main uncertainty they generate is driven by the processing times ξ. We tackle this problem by following an asymptotic analysis that will allow us to exploit a more manageable structure through careful scaling of the different parameters and thus reducing the complexity of the dynamics while retaining the essential characteristics and specifically the right trade-offs of the original problem, leading to limiting results that would naturally be good candidates for equilibrium approximations of the original problem. A common scaling mode in the context of queueing theory is obtained by driving the system into so-called **heavy traffic** or balanced loading. It's worth noting that multiple heavy traffic regimes may exist, each yielding different outcomes depending on the specified conditions.

In our context, a natural scaling is the one obtained by letting the demand rate grows linearly in some parameter n: $\lambda^n = n\,\lambda$. We parameterize all the other relevant quantities by n, in particular S^n, N^n and $\hat{c}^n$. Regarding the latter quantity, we assume that

$$\hat{c}^n/\sqrt{n} \Rightarrow \hat{c},$$

as $n \to \infty$, where $\hat{c}$ is a random variable with known distribution. Our analysis relies on the following Central-limit-theorem-like result for the number of items at the supply function level. Recall that $\zeta = (\zeta_i : i \geq 1)$ which is the sequence of expected value of the processing times at each supplier conditioned on the value of the capacity selected by the supplier. It is practically, the service time sequence of the infinite server queue that make-up the supply function.

Proposition 1 *When $n \to \infty$,*

$$\frac{N^n_s - n\,\mathbb{E}\left[\zeta_1\right]}{\sqrt{n}} \Rightarrow Z \sim \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \int_0^\infty F(x)\bar{F}(x)dx$ with $\bar{F}(x) = \mathbb{P}(\zeta_1 > x)\mathbb{E}_\Gamma\mathbb{P}(\xi_1 > x|\hat{\mu})$.

Lemma 2 *Assume that all the processing times are exponentially distributed with rate μ_k, $\xi_k \sim \exp(\mu_k)$, then*

$$\sigma^2(\boldsymbol{\mu}) = \mathbb{E}[\frac{1}{\hat{\mu}'}] - \mathbb{E}[\frac{1}{\hat{\mu}+\hat{\mu}'}] = \mathbb{E}_{\Gamma,\xi}\max\{0, \xi - \xi'\}.$$

Moreover, the following critical bound holds no matter the distribution Γ,

$$\sigma^2(\boldsymbol{\mu}) \geq \frac{1}{2\mathbb{E}_\Gamma\hat{\mu}}. \tag{3.4}$$

As for the number of items at the retailer's, we base our analysis on recent results of [21]. Under some minimal moments assumption on the perturbations ξ (light tailed or at least not too heavy-tailed) [21] show that under heavy traffic, $\rho \to 1$,

$$\frac{\log\log\left(\frac{1}{1-\rho}\right)}{\log\left(\frac{1}{1-\rho}\right)} W^\rho(\infty) \Rightarrow \frac{1}{\beta},$$

where $\beta > 1$. Given that $\mathrm{N}_\mathrm{r} = \lceil W^\rho_r(t)\rceil$ and by scaling ρ by n in our case, so that $\rho^n \to 1$ as $n \to \infty$, we obtain based on the results of [21] for an $S/D/1$, that

$$\frac{\mathrm{N}^n_\mathrm{r}}{\sqrt{n}} \to 0 \ a.s.$$

as $n \to \infty$.

Based on the above result, we therefore have that $\mathrm{N}^n \approx \mathrm{N}^n_\mathrm{s}$ for n large. For this reason, we drop N^n_r from our analysis for rest of this work i.e., the number of items in the system is equivalent to the number of items in the supply function.

Chapter 4

Centralized And Decentralized Solutions

4.1 Centralized Setting

We start our analysis by considering the centralized system, where the entire supply chain (retailer and supply function) is governed by one decision maker, namely the retailer. This vertically integrated view of the supply chain is interesting by itself but more importantly establishes a benchmark for evaluating the efficiency of the fully distributed system introduced in the previous chapter. In this scenario, a single decision maker (namely the retailer) has full control over the entire supply chain and would therefore simultaneously optimize for both the base-stock level and the capacity of each supplier in the objective to minimize the total cost of the supply chain.

Following the asymptotic formulation set-up discussed previously, we write the scaled supply chain aggregate cost denoted by $\mathrm{TC}^n(S^n, \boldsymbol{\mu})$:

$$\begin{aligned} \frac{\mathrm{TC}^n(S^n, \boldsymbol{\mu})}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \left[\mathrm{TC}^n_{\mathrm{r}}(S^n; \boldsymbol{\mu}) + \mathrm{TC}^n_{\mathrm{s}}(\boldsymbol{\mu}; S^n)\right] \\ &= \frac{1}{\sqrt{n}} \left[h\, \mathbb{E}[S^n - N^n]^+ + b\, \mathbb{E}[N^n - S^n]^+ + \mathbb{E}[\hat{c}^n\, \hat{\mu}]\right]. \end{aligned} \tag{4.1}$$

We start by showing that for any value of S and $\boldsymbol{\mu}$, the total cost can be asymptotically well defined.

We denote by S^n a base-stock level for system n and define S_0^n such that

$$S^n = n\, \mathbb{E}\zeta + \sqrt{n}\, \sigma\, z_0^n$$

and impose that $z_0^n \to z_0$.

We write the total centralized cost as follows:

$$\begin{aligned}\frac{\mathrm{TC}^n(S^n;\boldsymbol{\mu})}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\left(h\,\mathbb{E}[S^n - n\,\mathbb{E}\zeta - (N^n - n\mathbb{E}\zeta)]^+ + b\,\mathbb{E}[N^n - n\mathbb{E}\zeta - (S_c^n - n\mathbb{E}\zeta)]^+\right) + \frac{1}{\sqrt{n}}\mathbb{E}[\hat{c}^n \\ &= h\sigma\,\mathbb{E}[z^n - X^n]^+ + b\,\sigma\,\mathbb{E}[X^n - z_0^n]^+ + \frac{1}{\sqrt{n}}\mathbb{E}[\hat{c}^n\,\hat{\mu}] \\ &\Rightarrow \widetilde{\mathrm{TC}}(z_0;\boldsymbol{\mu}) \triangleq h\sigma\,\mathbb{E}[z_0 - Z]^+ + b\,\sigma\,\mathbb{E}[Z - z_0]^+ + \mathbb{E}[\hat{c}\,\hat{\mu}]\end{aligned}$$

as $n \to \infty$, where we set $X^n = \frac{N^n - n\mathbb{E}[\zeta]}{\sigma\sqrt{n}}$ and used Proposition 1 and the Continuous Mapping Theorem, to reach the above (weak) limit.

The optimization problem that the centralized decision maker is facing is now reduced at the limit to:

$$\min_{z_c,\boldsymbol{\mu}} \widetilde{\mathrm{TC}}(z_c,\boldsymbol{\mu}), \tag{4.2}$$

We denote by $S^{*,n}$ the optimal base-stock level for system n and define $S_0^{*,n}$ such that

$$S^{*,n} = n\,\mathbb{E}\zeta + \sqrt{n}\,S_0^{*,n}.$$

Proposition 2 *When $n \to \infty$, then*

$$S_0^{*,n} \to S_0^* = \sigma(\boldsymbol{\mu})\,z_0,$$

where σ is defined in Proposition 1, and $z_0 = \Phi^{-1}(\frac{b}{b+h})$, with Φ being the standard normal cumulative distribution.

Proof: Recall that

$$\frac{\mathrm{TC}^n(S_0^n;\boldsymbol{\mu})}{\sqrt{n}} = h\sigma\,\mathbb{E}[S_0^n/\sigma - X^n]^+ + b\,\sigma\,\mathbb{E}[X^n - S_0^n/\sigma]^+ + \mathbb{E}[\hat{c}^n/\sqrt{n}\,\hat{\mu}].$$

With respect to the optimization in S_0^n, this is a Newsvendor-like formulation with a constant terms $\mathbb{E}[(\hat{c}^n/\sqrt{n})\,\hat{\mu}]$ and σ, and hence the optimal centralized stock, $S_0^{*,n}/\sigma = (G^n)^{-1}(z_0)$ where G^n is the cumulative distribution of X^n/σ. Again, by Proposition 1 and the Continuous Mapping Theorem, we have that $G^n(\cdot) \to \Phi(\cdot)$ which completes the proof. ■

The previous result shows that the optimal base-stock level i.e., the one that minimizes the total supply chain cost is given by

$$S^{*,n} \approx n\,\mathbb{E}\zeta + \sqrt{n}\,S_0^*.$$

Interestingly, in this two stage supply chain yet with this multi-layer set of suppliers, we retrieve a Newsvendor-like solution for the retailer.

As for the suppliers' capacity, we denote by $\boldsymbol{\mu}^*$ the optimal capacity in the centralized case that minimizes the limiting total supply chain cost given in (4.2). In order to simplify the analysis, we assume that the perturbation sequence ξ_1 follows an exponential distribution each with some rate μ (a realization of $\hat{\mu}$, i.e. depending to which supplier the order is allocated). The problem remains complex given the infinite dimensional nature of the optimization. The next result is fundamental in solving this problem, as it allows one to reduce it to a two-dimensional optimization problem. We do that, by first recognizing that any centralized solution (4.2), defines an empirical distribution of $\hat{\mu}$ that we assume to admit at the limit a continuous distribution, and then we show that the cost minimization is equivalent to the one where the limiting distribution of $\hat{\mu}$ is reduced to a uniform distribution. It is worth noting that the result is obtained with no restriction on the distribution of $\hat{c}$.

Proposition 3 *Suppose $\hat{c}$ is a random variable with some known distribution and let $\boldsymbol{\mu}^*$ be the solution of (4.2). Then, there exist positive real numbers, a, b, with $a < b$, as well as $\tilde{\mu}^* \sim \mathcal{U}(a, b)$ such that $\widetilde{TC}(z_0, \boldsymbol{\mu}^*) = \widetilde{TC}(z_0, \tilde{\mu}^*)$.*

Proof:

Recall from Chapter (3) that:

$$\sigma^2(\tilde{\boldsymbol{\mu}}) = \mathbb{E}[\frac{1}{\tilde{\mu}}] - \mathbb{E}[\frac{1}{\tilde{\mu} + \tilde{\mu}'}]$$

and let

$$r(\tilde{\boldsymbol{\mu}}) = \mathbb{E}[\hat{c}\tilde{\boldsymbol{\mu}}].$$

Suppose that the retailer identified the optimal capacity levels for each supplier, and let $\boldsymbol{\mu}^*$ solution of $\mathrm{TC}(S, \boldsymbol{\mu})$. Denote by σ^{*2} and r^* values of σ^2 and r that correspond to this optimal value. We would like to show next that we can still generate the same values of σ^{*2} and r^* if we restrict the solution μ^* to be generated from a uniform distribution. Such claim would mean that it is enough for the retailer to set the capacities of the retailer in a way that they generate a uniform distribution.

Suppose that $\hat{c}$ is an r.v. with distribution F_c and WLOG assume that $\mathbb{E}\hat{c} = 1/2$. We can write $\hat{c} = F_c^{-1}(U)$ where $U \sim \mathcal{U}(0, 1)$ and F_c^{-1} is the inverse cumulative distribution which is also an increasing function. Now, suppose that the optimal equilibrium distribution Γ has been obtained which is basically the distribution of $\hat{\mu}$. Knowing Γ fixes also the value of σ^2 as well. As for r, given that the retailer is minimizing $\mathbb{E}\hat{c}\hat{\mu}$, each supplier can be parameterized by a value of $\hat{c}$ and the retailer will be better off allocating larger values of $\hat{c}$ to smaller values of $\hat{\mu}$. To do that, we can write $\mu(U)$, as a decreasing linear function that assigns a value of μ to each value of U. we write $\mu^*(U) = pU + q$. Since $U \sim \mathcal{U}(0, 1)$ and $\hat{\mu} \sim \mathcal{U}(a, b)$, then $\mu^*(0) = b$ and $\mu^*(1) = a$. Hence, we get that $p = (a - b)$ and $q = b$. It must be that $\mu^*(U) = b - (b - a)U$. Now we write the system of equations that need to be satisfied for such μ to be optimal,

$$r^* = \mathbb{E}[\hat{c}\boldsymbol{\mu}] = \mathbb{E}_U[F_c^{-1}(U)\mu(U)] = \int_0^1 F_c^{-1}(t)\mu(t)dt = b\mathbb{E}\hat{c} - (b-a)\int_0^1 tF_c^{-1}(t) = \frac{b}{2} - (b-a)A,$$

where $A = \int_0^1 tF_c^{-1}(t)dt > 0$. Moreover, consider $\tilde{\mu}, \tilde{\mu}' \sim \mathcal{U}(a, b)$.

$$
\begin{aligned}
\sigma^2(\tilde{\boldsymbol{\mu}}) = \mathbb{E}[\frac{1}{\tilde{\mu}}] - \mathbb{E}[\frac{1}{\tilde{\mu}+\tilde{\mu}'}] &= \frac{1}{b-a}\int_a^b \frac{1}{\tilde{\mu}}d\tilde{\mu} + \frac{1}{(b-a)^2}\int_a^b\int_a^b \frac{1}{\tilde{\mu}+\tilde{\mu}'}d\tilde{\mu}d\tilde{\mu}' \\
&= \frac{\ln(b)-\ln(a)}{b-a} - \frac{1}{(b-a)^2}\int_a^b \ln(b+\hat{\mu}') - \ln(a+\hat{\mu}')d\hat{\mu}' \\
&= \frac{\ln(b)-\ln(a)}{b-a} - \frac{(2b\ln(2b) - 2(a+b)\ln(a+b) + 2a\ln(2a))}{(b-a)^2} \\
&= \frac{\ln(b)-\ln(a)}{b-a} - \frac{2}{(b-a)^2}\left(b\ln(2b) + a\ln(2a) - (a+b)\ln(a+b)\right) \\
&= \frac{(b-a)(\ln(a)-\ln(b)) - 2b\ln(2b) - 2a\ln(2a) + 2(a+b)\ln(a+b)}{(b-a)^2} \\
&= \frac{-(a+b)(\ln(a)+\ln(b)) - 2(a+b)\ln(2) + 2(a+b)\ln(a+b)}{(b-a)^2} \\
&= \frac{-(a+b)\left(\ln(b)+\ln(a)+2\ln(2)-2\ln(a+b)\right)}{(b-a)^2}.
\end{aligned}
$$

We now need to show that the system

$$
\begin{cases}
\frac{-(a+b)(\ln(b)+\ln(a)+2\ln(2)-2\ln(a+b))}{(b-a)^2} = \sigma^{*2}, \\
\frac{b}{2} - (b-a)A = r^*.
\end{cases}
$$

has a unique solution (a^*, b^*).

Using $r^* = \frac{b}{2} - (b-a)A$, we write

- $a = \frac{r^*}{A} - (\frac{1}{2A} - 1)b$,
- $b - a = \frac{b}{2A} - \frac{r^*}{A}$,
- $a + b = \frac{r^*}{A} - \left(\frac{1}{2A} - 2\right)b$.

Given that a is strictly smaller than b, we should note that $r^* < b/2$ is a necessary condition. The equations above will then allow us to write σ^{*2} as a function of b, given by

$$
\sigma^{2*}(b) = \frac{-(\frac{r^*}{A} - \left(\frac{1}{2A}-2\right)b)\left(\ln(b) + \ln(\frac{r^*}{A} - (\frac{1}{2A}-1)b) + 2\ln(2) - 2\ln(\frac{r^*}{A} - \left(\frac{1}{2A}-2\right)b)\right)}{(\frac{b}{2A} - \frac{r^*}{A})^2}.
$$

Solving for b, is equivalent to solving for the solution of the following equality

$$
\ln\frac{4b\left(\frac{r^*}{A} - b(\frac{1}{2A}-1)\right)}{\left(\frac{r}{A} - (\frac{1}{2A}-2)b\right)^2} = \frac{-\left(\frac{b}{2A} - \frac{r^*}{A}\right)^2\sigma^{2*}}{\frac{r^*}{A} - \left(\frac{1}{2A}-2\right)b}. \tag{4.3}
$$

For $b < 2r$, there is no solution as the condition would not be satisfied. As for $b = 2r$, both the LHS and RHS are equal to 0. Their derivatives with respect to b are respectfully given by

$$\frac{-2r^*(b-2r^*)}{b\left((2A-1)b+2r^*\right)\left((4A-1)b+2r^*\right)} \quad \text{and} \quad \frac{-\left(b-2r\right)\left(\left(4A-1\right)b+\left(8A+2\right)r\right)\sigma^{*2}}{2A\left(\left(4A-1\right)b+2r\right)^2}.$$

We define the first one by $f(b)$ and the second by $g(b)$. Given that $b > 2r^*$ and $A > 0$, then both the RHS and LHS have a negative derivative and thus are decreasing in b. Moreover, we can easily check that for $b = \frac{2r^*}{1-2A}$,the LHS goes to $-\infty$ while the RHS goes to a constant. To show that (4.3) has a solution we need to look at the derivatives and show that the RHS decreases faster. That is because both are strictly decreasing but the LHS goes to $-\infty$ at $b = \frac{2r*}{1-2A}$ while the RHS goes to a real valued constant. We first note that both derivatives are equal to 0 at $b = 2r^*$. Since both derivatives are equal to 0 at $b = 2r^*$, we let $\epsilon > 0$ small enough and compare the derivatives at $b = 2r^* + \epsilon$. Before comparing we compute $f(2r^* + \epsilon)$ and $g(2r^* + \epsilon)$.

$$f(2r^*+\epsilon) = \frac{-2r^*\epsilon}{b\left((2A-1)b+2r^*\right)\left((4A-1)b+2r^*\right)}.$$

Before proceeding, we look at the denominator and note that it is equal to

$$(2A-1)(4A-1)b^3 + 2r^*(2A-1)b^2 + 2r^*(4A-1)b^2 + b(2r^*)^2.$$

As $b = 2r^* + \epsilon$, we note that b^3 will be equal to $(2A-1)(4A-1)(2r^*)^3$ plus $o(\epsilon^3)$ with some constant as coefficient. Similarly for b^2, we get $o(\epsilon)^2$ and for b with $o(\epsilon)$. We disregard these terms as they are not relevant to our analysis. This is equivalent to solving

$$\begin{aligned}
&\frac{-2r^*\epsilon}{(2A-1)(4A-1)(2r^*)^3 + 2r^*(2A-1)(2r^*)^2 + 2r^*(4A-1)(2r^*)^2 + (2r^*)(2r^*)^2} \\
&= \frac{-\epsilon}{(2A-1)(4A-1)(2r^*)^2 + 2r^*(2r^*)\left(4A-1+2A-1\right) + (2r^*)^2} \\
&= \frac{-\epsilon}{(2r^*)^2\left((2A-1)(4A-1)+6A-2+1\right)} = \frac{-\epsilon}{(2r^*)^2\left(8A^2\right)}
\end{aligned}$$

As for the other function we have

$$g(2r^*+\epsilon) = \frac{-\epsilon\left((4A-1)b + 2(4A+1)r^*\right)\sigma^{*2}}{2A\left((4A-1)b+2r^*\right)^2}$$

By doing the same for both the numerator and denominator, it becomes equivalent to solving

$$\begin{aligned}
&\frac{-\epsilon\left((4A-1)(2r^*) + (4A+1)(2r^*)\right)\sigma^{*2}}{2A\left((4A-1)(2r^*)+2r^*\right)^2} \\
&= \frac{-\epsilon\left((2r^*)8A\right)\sigma^{*2}}{2A\left((2r^*)(4A)\right)^2} = \frac{-\epsilon\,\sigma^{*2}}{8A^2r^*}
\end{aligned}$$

For $|\frac{-\epsilon}{(2r^*)^2(8A^2)}| < |\frac{-\epsilon\sigma^{*2}}{8A^2r^*}|$ to hold, we need

$$\sigma^{2*} > \frac{1}{4r^*}.$$

Note also based on our discussion above $\hat{\mu}^*$ and $\hat{c}$ are negatively correlated so that

$$r^* \leq \mathbb{E}\hat{c}\,\mathbb{E}\hat{\mu} = \bar{\mu}/2.$$

We confirm the above condition in the case where $\hat{c} \sim \mathcal{U}(0,1)$. It is not hard to show in this case that $A = 1/3$, which implies that $f(b) = \frac{18r^*(b-2r^*)}{b(b^2-36r^*)}$ and $g(b) = \frac{-9(b-2r^*)(b+14r^*)\sigma^{*2}}{2(b+6r^*)^2}$. Then, by evaluating these functions at $b = 2r^* + \epsilon$, in absolute value we get that $|f(2r^*+\epsilon)| < |g(2r^*+\epsilon)|$ only if $\sigma^{*2} > \frac{1}{4r^*}$. In conclusion, the proof of the theorem is completed once we show that

$$\sigma^{*2} > \frac{1}{4r^*},$$

and a sufficient condition is that for any distribution of μ, we have that

$$\sigma^2(\mu) \geq \frac{1}{2\,\bar{\mu}} \tag{4.4}$$

which is guaranteed by Lemma 2. ■

The result above characterizes the general solution for any distribution of $\hat{c}$. A special case for the centralized solution would be if $\hat{c}$ was assumed to be constant, i.e. $\hat{c} = c$. The Corollary below characterizes its solution.

Corollary 1 *If c is constant i.e. $\hat{c} = c$ and the processing times are assumed to be exponential, i.e, $\xi_k \sim \exp(\mu_k)$, then, for any supplier k, at the limit as $n \to \infty$,*

$$\mu^* = \frac{1}{2}\left(\frac{((b+h)\,\mathbb{E}[Z-z_0]^+ + h\,z_0)}{c}\right)^{2/3} = \frac{1}{2}\left(\frac{\Psi(z_0)}{c}\right)^{2/3},$$

where Ψ is the optimal standardized cost with $\Psi(z_0) = h\,\mathbb{E}[z_0 - Z]^+ + b\,\mathbb{E}[Z - z_0]^+$. Finally, the limiting total expected supply chain cost is given by:

$$\widetilde{TC}^{\infty}(\boldsymbol{\mu}^*, z_0) = \Psi(z_0)\,\sigma(\boldsymbol{\mu}^*) + c\mu^* = \frac{3}{2}\,\Psi(z_0)^{2/3}\,c^{1/3}.$$

Proof: We give here a direct proof of the result.

Step 1. Obtain a lower bound for the total centralized cost. Define first the function

$$\Psi(z) = \left((h+b)\mathbb{E}[Z-z]^+ + hz\right).$$

Using the fact that $\mathbb{E}[X - z_0] = \mathbb{E}[X - z_0]^+ - \mathbb{E}[z_0 - X]^+$, we rewrite the limit of (4.1) as $n \to \infty$,

$$\widetilde{\mathrm{TC}}^{\infty}(\boldsymbol{\mu}) \triangleq \sigma\Psi(z_0) + c\bar{\mu}. \tag{4.5}$$

By using a similar approach as in the proof of Lemma 2, we know that $\sigma^2(\boldsymbol{\mu}) \geq \mathbb{E}_\Gamma \frac{1}{2\hat{\mu}}$. In fact we follow the same method but skip the use of the Jensen Inequality, we reach our desired result. Applying this to equation (4.5) leads to the following inequality

$$\widetilde{\mathrm{TC}}(\boldsymbol{\mu}) \geq \left(\mathbb{E}_\Gamma \frac{1}{2\hat{\mu}}\right)^{1/2} \Psi(z_0) + c\mathbb{E}_\Gamma \hat{\mu}.$$

By Jensen's inequality, $\left(\mathbb{E}_\Gamma \frac{1}{2\hat{\mu}}\right)^{1/2} \geq \mathbb{E}_\Gamma \frac{1}{(2\hat{\mu})^{1/2}}$. Then,

$$\widetilde{\mathrm{TC}}(\boldsymbol{\mu}) \geq \mathbb{E}_\Gamma \frac{1}{\sqrt{2\hat{\mu}}} \Psi(z_0) + c\mathbb{E}_\Gamma \hat{\mu} = \mathbb{E}_\Gamma \left(\frac{1}{\sqrt{2\hat{\mu}}} \Psi(z_0) + c\hat{\mu}\right)$$

From the Mean Value Theorem, $\exists\, \mu_0$ such that $\mathbb{E}_\Gamma \left(\frac{1}{2\hat{\mu}} \Psi(z_0) + c\hat{\mu}\right) = \frac{1}{2\mu_0} \Psi(z_0) + c\mu_0$. Define the function $g(\mu_0) = \frac{1}{\sqrt{2\mu_0}} \Psi(z_0) + c\,\mu_0$. It is easy to show that g is convex as the second derivative, $g''(\mu_0) = \frac{3\Psi(z_0)}{(2\mu_0)^{5/2}} > 0$. We solve for the FOC and obtain that $g'(\mu_0) = \frac{-\Psi(z_0)}{(2\mu_0)^{3/2}} + c = 0$. So that g admits a minimum $g^* = g(\mu_0^*)$ with $\mu_0^* = \frac{1}{2}\left(\frac{\Psi(z_0)}{c}\right)^{2/3}$.

We conclude that

$$\widetilde{\mathrm{TC}}(\boldsymbol{\mu}) \geq g^*.$$

Step 2. Show that the minimum of g is also a minimum of $\widetilde{\mathrm{TC}}^\infty$. In other words, set $\boldsymbol{\mu}_0^*$ such that $\mu_k = \mu_0^*$ for all k and show that $\tilde{TC}^\infty(\boldsymbol{\mu}_0^*) = g^*$.
First, note that for a constant r.v. $\hat{X}$, $\mathbb{E}[\frac{1}{\hat{X}}] = \frac{1}{X}$ and $\mathbb{E}[\frac{\hat{X}}{\hat{X}+\hat{X}} = \frac{1}{2X}$. Hence, for $\boldsymbol{\mu} = \boldsymbol{\mu}_0^*$,

$$\sigma = \frac{1}{\sqrt{2\mu_0^*}} = \left(\frac{\Psi(z_0)}{c}\right)^{-1/3}.$$

Hence,

$$\widetilde{\mathrm{TC}}(\boldsymbol{\mu}_0^*) = \left(\frac{\Psi(z_0)}{c}\right)^{-1/3} \Psi(z_0) + c\mu_0^* = \frac{3}{2}\Psi(z_0)^{2/3} c^{1/3}.$$

This is exactly equal to $g*$. Since g^* is a global minimum and $\tilde{TC}(\boldsymbol{\mu_0^*}) = g^*$, then μ_0^* is a global minimum of $\widetilde{\mathrm{TC}}^\infty(\boldsymbol{\mu})$. Hence, $\tilde{TC}(\boldsymbol{\mu})$ admits a global minimum at

$$\mu^* \triangleq \frac{1}{2}\left(\frac{\Psi(z_0)}{c}\right)^{2/3}$$

where $\Psi(z_0) = (b+h)\mathbb{E}[Z - z_0]^+ + hz_0$. ■

4.1.1 *Discussion and Numerical Results*

Before moving on to the decentralized solutions, we discuss the variation of the optimal solutions with respect to a change in the parameters; b, h and $\hat{c}$. We introduce a lemma that will be used throughout this work.

Lemma 3 *A closed form for* $\mathbb{E}[Z-z_0]^+$ *using the truncated expectation of a normal random variable is given by*

$$\mathbb{E}[Z-z_0]^+ = \left(-z_0 + \frac{\varphi(z_0)}{\bar{\Phi}(z_0)}\right)\bar{\Phi}(z_0) = -z_0\bar{\Phi}(z_0) + \varphi(z_0)$$

We note that this discussion is performed for $\hat{c} = c$. It's clear that $S^*_{0,c}$ increases with the capacity cost c and therefore, the optimal base-stock level increases with c.

As for the holding and backlog costs, it is much harder to asses the trend of the base-stock level with respect to these parameters since $S^*_0 = z_0\,\sigma^*$ where both components depend on h and b. For that we look at

$$\frac{\partial S^*_0}{\partial h} = c^{1/3}\left(-\frac{b}{(b+h)^{7/3}\varphi(z_0)^{4/3}} - \frac{1}{3}z_0\frac{(\varphi(z_0) + z_0\Phi(z_0))}{((b+h)\varphi(z_0))^{4/3}}\right),$$

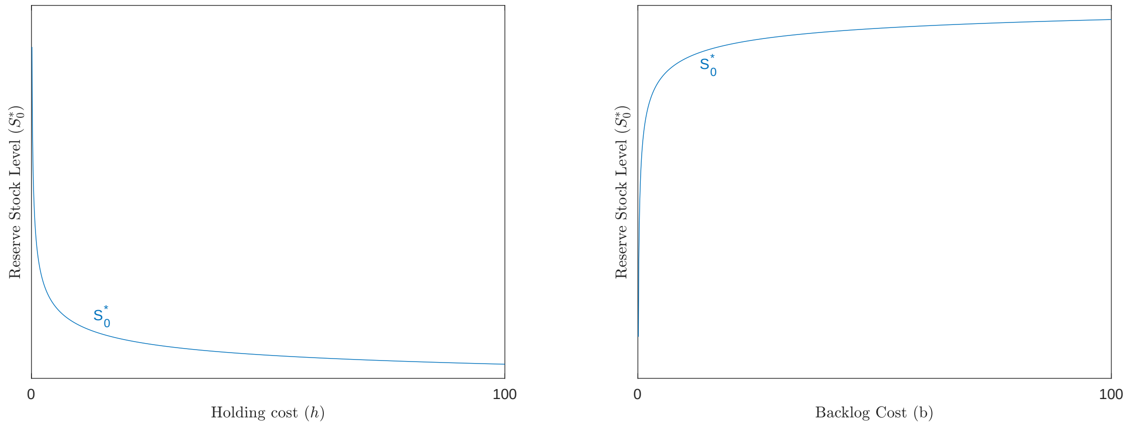
and

$$\frac{\partial S^*_0}{\partial b} = c^{1/3}\left(\frac{h}{(b+h)^{7/3}\varphi(z_0)^{4/3}} - \frac{1}{3}z_0\frac{\left(\varphi(z_0) - z_0\bar{\Phi}(z_0)\right)}{((b+h)\varphi(z_0))^{4/3}}\right).$$

We note that all elements are positive except for z_0. For that we split the problem into two cases, starting with the obvious ones: if $z_0 > 0$ then it is clear that $\frac{\partial S^*_0}{\partial h} < 0$. and if $z_0 < 0$ then $\frac{\partial S^*_0}{\partial b} > 0$. As for the second case, and looking at the derivative of with respect to h, if $z_0 < 0$, then given that $\varphi(x) = \varphi(-x)$ we know that the first fraction is negative. For the second we use the fact that $\Phi(-x) = \bar{\Phi}(x)$ to write $\varphi(-z_0) - z_0\Phi(-z_0) = \varphi(z_0) - z_0\bar{\Phi}(z_0)$. Through lemma 3, $\mathbb{E}[Z-z_0]^+ = \varphi(z_0) - \bar{\Phi}(z_0)z_0 > 0$. The denominator of both fractions are positive. We need to show now that the first fraction dominate the second. One could look at the limits of z_0 and see that the first one will be larger but they need to account for all cases.

Similarly for b, after following the same procedure and reaching similar conclusions, we look for a numerical comparison. We note that what makes z_0 negative or positive is the ratio of $b/(b+h)$. For z_0 to be positive, we need $\Phi^{-1}\left(\frac{b}{b+h}\right) > 0$, i.e. we need $b/(b+h) > 1/2$. This means that for $b > h$, then $z_0 > 0$ and vice versa. The graphs below show that S^*_0 is strictly decreasing in h and increasing in b as they account for both cases of z_0. Hence, the optimal base-stock level increases with the backlog cost and decreases with the holding cost.

Figure 4.1: Optimal safety-stock levels



As for the optimal capacity level, it is also clear that it decreases with c. We also note that μ^* can be seen as a function of the limit of the optimal total centralized cost. In fact, we re-write

$$\mu^* = \frac{1}{3}\frac{\widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu^*})}{c}.$$

It is interesting to see that $\mu^* \propto \widetilde{TC}(z_0, \boldsymbol{\mu^*})$ and that choosing a higher capacity level will increase the total cost at a $3\,c$ rate. To analyse the effect of h and b, we first look at the derivative of $\tilde{TC}(z_0, \boldsymbol{\mu^*})$ with respect to both parameters;

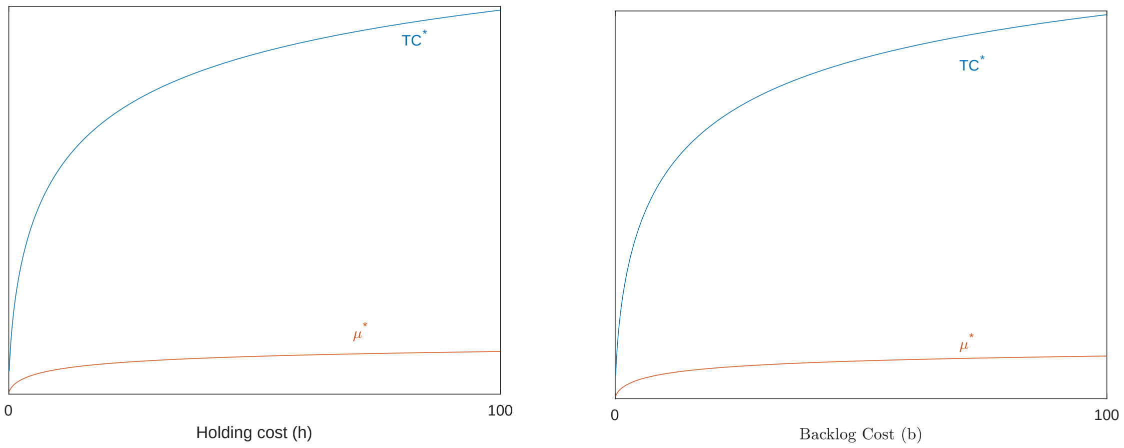
$$\frac{\partial \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu^*})}{\partial h} = \left(\frac{c}{(b+h)\varphi(z_0)}\right)^{1/3} \left(\varphi(z_0) + z_0\Phi(z_0)\right),$$

and

$$\frac{\partial \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu^*})}{\partial h} = \left(\frac{c}{(b+h)\varphi(z_0)}\right)^{1/3} \left(\varphi(z_0) - z_0\bar{\Phi}(z_0)\right).$$

Using similar calculations and plotting as for S_0^*, we conclude that $\frac{\partial \widetilde{\mathrm{TC}}(z_0,\boldsymbol{\mu^*})}{\partial h} > 0$ and $\frac{\partial \widetilde{\mathrm{TC}}(z_0,\boldsymbol{\mu^*})}{\partial b} > 0$ for both $z_0 \leq 0\,(h > b)$ and $z_0 \geq 0\,(h < b)$.

Figure 4.2: Optimal limiting centralized costs



As expected, the limit of the total optimal centralized cost will increase with both the holding and backlog costs. As a result, the optimal capacity level will increase with h and b. It is also obvious that this limit will increase with the capacity cost. Finally, it is natural to note that the backlog split α doesn't affect the centralized solution since the payments between agents do not affect the centralized cost.

4.2 Decentralized Setting

We consider two scenarios for the decentralized setting. The first one assumes that the suppliers are centralized in the sense that there is one decision maker for the supply function. The second scenario is the fully decentralized case where each supplier selects independently his capacity. In both cases we proceed to find the Nash equilibrium between the agents. In other words, the retailer will find his optimal base-stock level, assuming that the suppliers (either as a supply function or individually) will minimize their cost by finding their optimal capacity.

The scaled cost of the retailer is given by

$$\frac{\mathrm{TC}_{\mathrm{r}}^n(S^n;\boldsymbol{\mu})}{\sqrt{n}} = \frac{1}{\sqrt{n}}\left(h\mathbb{E}[S^n - N^n]^+ + \alpha\, b\mathbb{E}[N^n - S^n]^+\right)$$

as $n \to \infty$.

We undertake a similar asymptotic analysis as above and write the optimal base stock level for system n, and define $S_{0,d}$ such that

$$S_d^n = n\mathbb{E}\zeta + \sqrt{n}\sigma z_\alpha^n,$$

and impose $z_\alpha^n \to z_\alpha$.

Similarly to the centralized system, we use Proposition 1 and the continuous Mapping theorem to write

$$\frac{\mathrm{TC}_{\mathrm{r}}^n(S^n;\boldsymbol{\mu})}{\sqrt{n}} \Rightarrow \widetilde{\mathrm{TC}}_{\mathrm{r}}(z;\boldsymbol{\mu}) \triangleq h\sigma\mathbb{E}[z_\alpha - Z]^+ + b\sigma\mathbb{E}[Z - z_\alpha]^+,$$

as $n \to \infty$ The optimization problem that the retailers solving is now reduced to

$$\min_{z_\alpha} \widetilde{\mathrm{TC}}_{\mathrm{r}}(z;\boldsymbol{\mu})$$

Proposition 4 *When $n \to \infty$,*

$$S_{0,d}^{*,n} \to S_{0,d}^* = \sigma(\boldsymbol{\mu})\, z_\alpha$$

where σ is defined in Proposition 1, and $z_\alpha = \Phi^{-1}(\frac{\alpha\, b}{\alpha\, b+h})$, with Φ being the standard normal cumulative distribution.

The proof follows similar steps than the one of Proposition 2 and will be skipped.

We note that σ is dependent on the optimal capacity levels of the suppliers that will be found below. The previous result shows that the optimal base-stock level i.e., the one that minimizes the total cost of the retailer is given by

$$S_d^{*,n} \approx n\mathbb{E}[\zeta] + \sqrt{n}S_{0,d}^*.$$

4.2.1 *Centralized Suppliers*

In this subsection we consider a retailer replenishing orders from one supply function represented by a large number of suppliers, and parameterized by $\bar{\mu}$. The agents will simultaneously find their optimal decision variables. Based on Proposition 4, the optimal base-stock level of the retailer in a centralized supplier case is given by $S^{*,n}_{cs} \approx n\mathbb{E}\zeta + \sqrt{n}S^*_{0,cs}$, where

$$S^{*,n}_{0,cs} \to S^*_{0,cs} = \sigma(\boldsymbol{\mu}^*_{cs})\, z_\alpha.$$

And where $\boldsymbol{\mu}^*_{cs}$ is the optimal capacity that minimizes the limiting total cost of the supply function. Again, for simplicity we assume that ξ_1 follows an exponential distribution with rate μ_{cs}.

The scaled cost of the supply function is given by

$$\begin{aligned} \frac{\mathrm{TC}^n_{\mathrm{s}}(\boldsymbol{\mu}; S^n)}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \left[(1-\alpha)\, b\, \mathbb{E}[N^n - S^n]^+ + \mathbb{E}[\hat{c}^n\, \hat{\mu}]\right] \\ &\Rightarrow \widetilde{\mathrm{TC}}_{\mathrm{s}}^{\infty}(\boldsymbol{\mu}; z_\alpha) \triangleq (1-\alpha)\, b\, \mathbb{E}[Z - z_\alpha]^+ + \mathbb{E}[\hat{c}\hat{\mu}] \end{aligned} \tag{4.6}$$

The optimization problem of the supply function is then reduced to

$$\min_{\boldsymbol{\mu}} \widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z_\alpha)$$

Proposition 5 *Given that $\hat{c}$ is a random variable with some known distribution, let $\boldsymbol{\mu}^*_c$ be the solution of $\widetilde{TC}_s^{\infty}(\boldsymbol{\mu}; z_\alpha)$. Then, there exist a, b and $\tilde{\mu}^*_{cs} \sim \mathcal{U}(a,b)$ such that $\widetilde{TC}_s^{\infty}(\boldsymbol{\mu}^*_{cs}; z_\alpha) = \widetilde{TC}_s^{\infty}(\tilde{\mu}^*_{cs}; z_\alpha)$.*

We note that the proof of this proposition follows from the proof of the centralized solution, as their cost functions are quite similar with only the coefficient of the truncated expected value being different.

Corollary 2 *If c is constant, i.e. $\hat{c} = c$ and the processing times are assumed to be exponential, i.e. $\xi_k \sim \exp(\mu_k)$, then for any supplier k, at the limit as $n \to \infty$,*

$$\mu^*_{cs} = \frac{1}{2}\left(\frac{(1-\alpha)b\mathbb{E}[Z - z_\alpha]^+}{c}\right)^{2/3} = \frac{1}{2}\left(\frac{\chi_\alpha}{c}\right)^{2/3}.$$

Finally, the limiting total expected supply chain cost is given by:

$$\widetilde{TC}_d^{\infty}(\boldsymbol{\mu}^*_{cs}, z_\alpha) = \Psi(z_\alpha)^{2/3} c^{1/3} \left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3} + \frac{1}{2}\left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right).$$

The proof of this corollary follows directly from the proof of Corollary 1. It is sufficient to replace $\Psi(z_0)$ with χ_α in the limiting cost to reach the optimal capacity of the supply chain.

4.2.2 *Decentralized Suppliers*

We finally tackle the most relevant case in our analysis of distributed system namely where each supplier, characterized by a capacity cost c_k (drawn from $\hat{c}$, a random variable with some known distribution) decides the capacity level μ_k that minimizes the individual cost $\mathrm{TC}_{\mathrm{s}}^{k}$. In this context we proceed to find a Nash equilibrium for all agents. In a such a setting, the retailer chooses the base-stock level S that minimizes TC_{r} while accounting for the suppliers choice of capacity levels $\hat{\mu}$ and vice versa. Similarly to the subsection above, the optimal base-stock level of the retailer in a centralized supplier case, is given by $S_{ds}^{*,n} \approx n\mathbb{E}\zeta + \sqrt{n}S_{0,ds}^{*}$, where

$$S_{0,ds}^{*,n} \to S_{0,ds}^{*} = \sigma(\boldsymbol{\mu}_{ds}^{*})\, z_{\alpha}.$$

And where $\boldsymbol{\mu}_{ds}^{*}$ is the vector of optimal capacities μ_k^* that minimizes the limiting total cost of each supplier. The scaled cost of supplier k is given by

$$\begin{aligned}\frac{\mathrm{TC}_{\mathrm{s}}^{k,n}(\mu_k^n; S^n|\Gamma)}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\left((1-\alpha)\, b\, \mathbb{E}[N-S]^{+}\left(\frac{1/\mu_k}{\mathbb{E}_{\Gamma}1/\hat{\mu}}\right) + c_k^n\, \mu_k\right)\\ &\Rightarrow \widetilde{\mathrm{TC}_{\mathrm{s}}}^{k}(\mu_k; z_\alpha|\Gamma) \triangleq (1-\alpha)\, b\, \mathbb{E}[Z-z_\alpha]^{+}\left(\frac{1/\mu_k}{\mathbb{E}_{\Gamma}1/\hat{\mu}}\right) + c_k\mu_k \quad (4.7)\end{aligned}$$

as $n \to \infty$. The optimization problem of the supplier k is then reduced to

$$\min_{\mu_k} \widetilde{\mathrm{TC}_{\mathrm{s}}}^{k}(\mu_k; z_\alpha|\Gamma)$$

As we have described in Chapter 3, we will be obtaining an approximation of this equilibrium through a Mean Field Nash Equilibrium approach applied to the limiting system (as we scale in n). The limiting mean field assumption allows to fix the distribution of the population (given the large number of players) i.e. of $\hat{\mu}$ whereby at equilibrium the solution μ_k of each supplier given the population distribution is aligned with that population distribution. Fixing the population distribution implies that the expected value over it will remain fixed. As such, given Γ, σ and $\mathbb{E}\mu^{-1}$ are fixed as the supplier is solving for the optimal μ_k^*. Applying this, we look for the optimal capacity level of each supplier:

Proposition 6 *If $\xi_1 \sim \exp(\mu)$, then*

$$\mu_k^* = \frac{\sqrt{\chi_\alpha}\left(\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1}\right)^{1/3}}{\left(\mathbb{E}\sqrt{\hat{\nu}}\right)^{2/3}}\frac{1}{\sqrt{c_k}}$$

with $\hat{\nu} = \hat{c}/\chi_\alpha$ where χ_α is the optimal decentralized standardized cost, such that $\chi_\alpha = (1-\alpha)b\mathbb{E}[Z-z_\alpha]^{+}$.

The limiting cost o fthe total supply chain is given by

$$\widetilde{TC_{d}}(z_\alpha, \boldsymbol{\mu}_{ds}^{*}) = \Psi(z_\alpha)\sigma(\boldsymbol{\mu}_{ds}^{*}) + \mathbb{E}[\hat{c}\hat{\mu}_{ds}^{*}] = \Psi(z_\alpha)\sigma(\boldsymbol{\mu}_{ds}^{*}) + \sqrt{\chi_\alpha}\sqrt{\frac{\sigma(\boldsymbol{\mu}_{ds}^{*})}{\mathbb{E}\mu_{ds}^{*,-1}}}\mathbb{E}\sqrt{\hat{c}}.$$

Where $\mathbb{E}\mu^{-1} = \frac{(\mathbb{E}\sqrt{\hat{\nu}})^2}{\sigma}$ *and* $\sigma(\boldsymbol{\mu}_{ds}^{*}) = \left(\mathbb{E}[\hat{\nu}^{1/2}](\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1})\right)^{1/3}$

Proof: We first look at the limit of equation (4.7),

$$\widetilde{\mathrm{TC}}_s^{\,k}(\mu_k) \triangleq \chi_\alpha \frac{\mu_k^{-1}}{\mathbb{E}_\Gamma \hat{\mu}^{-1}}\sigma + c_k\mu_k.$$

By taking solving the First Order Condition we get that

$$\hat{\mu} = \sqrt{\frac{\chi_\alpha \sigma/\mathbb{E}_\Gamma \hat{\mu}^{-1}}{\hat{c}}}.$$

Set $\hat{\nu} = \hat{c}/\chi_\alpha$. Then, by inverting and taking the average on both ends we conclude that

$$\mathbb{E}\mu^{-1} = \frac{(\mathbb{E}\sqrt{\hat{\nu}})^2}{\sigma}. \tag{4.8}$$

Hence,

$$\hat{\mu} = \frac{\sigma}{\sqrt{\hat{\nu}}\,\mathbb{E}\sqrt{\hat{\nu}}}. \tag{4.9}$$

We write $\hat{\mu} + \hat{\mu}' = \frac{\sigma}{\mathbb{E}\sqrt{\hat{\nu}}}(\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2})$, so that

$$\sigma^2 = \mathbb{E}[\frac{1}{\hat{\mu}}] - \mathbb{E}[\frac{1}{\hat{\mu} + \hat{\mu}'}] = \frac{\mathbb{E}\sqrt{\hat{\nu}}}{\sigma}(\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1}),$$

equivalently:

$$\sigma = \left(\mathbb{E}[\hat{\nu}^{1/2}](\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1})\right)^{1/3}. \tag{4.10}$$

By putting together equations (4.9) and (4.10), we obtain the unique equilibrium distribution Γ^* of $\hat{\mu}^*$. ■

Corollary 3 *If c is constant i.e. $\hat{c} = c$ and the processing times are assumed to be exponential, i.e, $\xi_k \sim \exp(\mu_k)$, then, for any supplier k, at the limit as $n \to \infty$,*

$$\mu_k^* = \frac{1}{2^{1/3}}\left(\frac{(1-\alpha)b\mathbb{E}[Z - z_\alpha]^+}{c}\right)^{2/3} = \frac{1}{2^{1/3}}\left(\frac{\chi_\alpha}{c}\right)^{2/3}.$$

Finally, the total supply chain cost is given by:

$$\begin{aligned}\widetilde{TC}_s^{\,k}(\boldsymbol{\mu}_{ds}^*; S_{ds}^*) = \Psi(z_\alpha)\sigma(\mu_k^*) + c\mu_k^* &= \Psi(z_\alpha)\frac{1}{2^{1/3}}\left(\frac{\chi_\alpha}{c}\right)^{-1/3} + \frac{c}{2^{1/3}}\left(\frac{\chi_\alpha}{c}\right)^{2/3} \\ &= \frac{c^{1/3}\Psi(z_\alpha)^{2/3}}{2^{1/3}}\left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3} + \left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right).\end{aligned}$$

4.3 Comparison of Solutions

In this section, we compare our optimal solutions and optimal limiting total supply chain costs. The following table 4.1 gives a brief summary of the optimal capacity level and base-stock level as well as the optimal total costs achieved in the three frameworks mentioned. We also mention that we are comparing the solutions where $\hat{c} = c$.

	Centralized system	Centralized suppliers	Decentralized suppliers
Capacity (μ^*)	$\frac{1}{2}\left(\frac{\Psi(z_0)}{c}\right)^{2/3}$	$\frac{1}{2}\left(\frac{\chi_\alpha}{c}\right)^{2/3}$	$\frac{1}{2^{1/3}}\left(\frac{\chi_\alpha}{c}\right)^{2/3}$
Safety-stock (S_0*)	$z_0\left(\frac{\Psi(z_0)}{c}\right)^{-1/3}$	$z_\alpha\left(\frac{\chi_\alpha}{c}\right)^{-1/3}$	$z_\alpha\frac{1}{2^{1/3}}\left(\frac{\chi_\alpha}{c}\right)^{-1/3}$
$\widetilde{TC}(S^*,\mu^*)$	$\frac{3}{2}\Psi(z_0)^{2/3}c^{1/3}$	$\Psi(z_\alpha)^{2/3}c^{1/3}\left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3}+\frac{1}{2}\left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right)$	$\frac{1}{2^{1/3}}\Psi(z_\alpha)^{2/3}c^{1/3}\left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3}+\left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right)$

Table 4.1: Comparison of Decision Variables and Total Costs

4.3.1 *Optimal Capacity Level*

Moving on to the capacity level chosen by the suppliers, it is easy to see that it is sufficient to compare ψ_1 andχ_α. This is not trivial as their ratio is highly dependent on α. For that we first look into χ_α and show that it is in fact a decreasing function in α. The derivative of χ_α is

$$\begin{aligned}\frac{\partial \chi_\alpha}{\partial \alpha} &= \frac{\partial}{\partial \alpha}(1-\alpha)b\mathbb{E}[Z-z_\alpha]^+ = -b\mathbb{E}[Z-z_\alpha]^+ + (1-\alpha)b\frac{\partial \mathbb{E}[Z-z_\alpha]^+}{\partial \alpha} \\ &= -\left(b\mathbb{E}[Z-z_\alpha]^+ + (\alpha)bz'_\alpha\bar{\Phi}(z_\alpha)\right)\end{aligned}$$

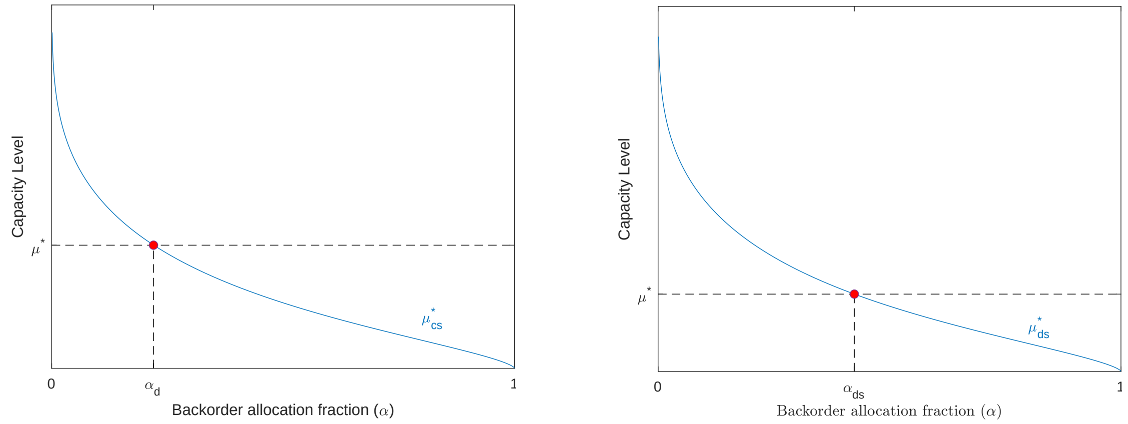
where $z'_\alpha = \frac{\partial z_\alpha}{\partial \alpha} = \frac{bh}{(\alpha b+h)^2\varphi(z_\alpha)} > 0$ Hence, as the derivative of χ_α is always negative, we now know that it is decreasing in α. Moreover, we note that as $\alpha \to 0$ and $\alpha \to 1$, $z_\alpha \to -\infty$ and $z_\alpha \to z_0$ respectively. Then, we conclude that χ_α is a decreasing function going from $+\infty$ to 0 with respect to $0 \le \alpha \le 1$.

Moving on, $\Psi(z_0)$ is a positive function that is independent of α. It then must intersect χ_α at a point for some $\alpha_d \in [0,1]$. This point is the solution to

$$\Psi(z_0) = \chi_\alpha \implies \frac{(h+b)\varphi(z_0)}{b} = (1-\alpha)\left(-z_\alpha\frac{h}{\alpha b+h} + \varphi(z_\alpha)\right)$$

It is hard to find α^* that solves this equation as it is mathematically intractable. We will resort to some computational method to learn more about the value of α^* but before that, we compare the centralized capacity with the two decentralized capacities. As χ_α is strictly decreasing in α and since $\Psi(z_0)$ is independent of it, we can easily see that $\Psi(z_0) < \chi_\alpha$ when $\alpha < \alpha^*$ and $\Psi(z_0) > \chi_\alpha$ when $\alpha > \alpha^*$.

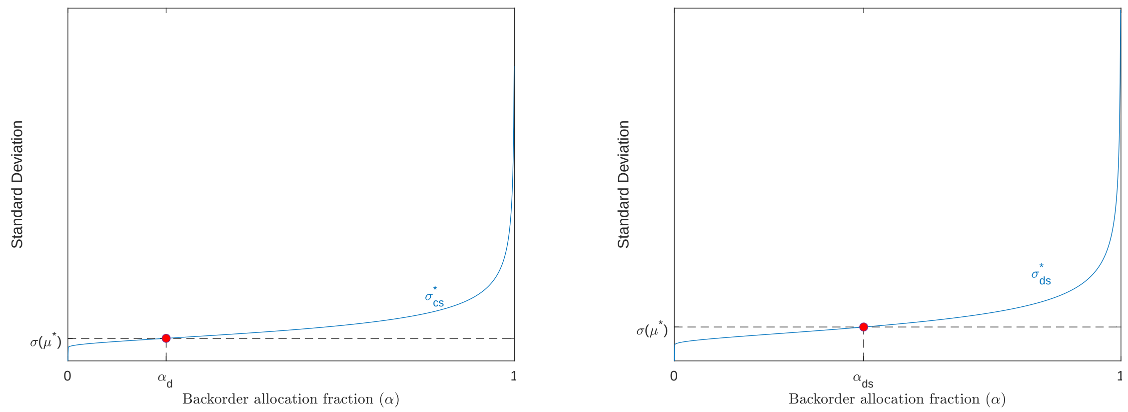
Figure 4.3: The Optimal Centralized and Decentralized Suppliers' Production Capacity μ^*_{cs} and μ^*_{ds} as a Function of the Retailer's Backorder Share α —The Centralized Solution is μ^*



It is clear that both capacities intersect the optimal capacity at some α. These plots also verify our result; for $\alpha < \alpha^*$, $\mu^* < \mu^*$ and $\alpha > \alpha^*$, $\mu^* > \mu^*$ for both capacities. For $h = 7, b = 10$ and $c = 3$, we get $\alpha_d = 0.22$ and $\alpha_{ds} = 0.4242$.

As for σ, we know that for μ constant, $\sigma = \frac{1}{\sqrt{2\mu}}$. Then, σ is a decreasing function is μ. We showed that both μ^* are decreasing functions in α. Hence, σ is an increasing function in α starting from 0 and going to $+\infty$. Then, we will get $\Psi(z_0) > \chi_\alpha$ when $\alpha < \alpha^*$ and $\Psi(z_0) < \chi_\alpha$ when $\alpha > \alpha^*$. We also plot $\sigma(\mu^*_{cs})$ and $\sigma(\mu^*_{ds})$ versus $\sigma(\mu^*_c)$ and check the values of α^*.

Figure 4.4: The Function σ Evaluated at Optimal Centralized and Decentralized Suppliers' Production Capacity μ^*_{cs} and μ^*_{ds} as a Function of the Retailer's Backorder Share α —The Centralized Solution is σ^*.



Again, it is clear that both σ's intersect the optimal one at some α. These plots also verify our result; for $\alpha < \alpha^*$, $\sigma(\mu^*) > \sigma(\mu^*_d)$ and $\alpha > \alpha^*$, $\sigma(\mu^*) < \sigma(\mu^*_d)$ for both capacities (μ^*_d is the optimal capacity for either ones of the decentralized solutions).

For $h = 7, b = 10$ and $c = 3$, we get $\alpha_d = 0.22$ and $\alpha_{ds} = 0.4242$. It is interesting to see that the α^* for the μ^*s and σ^* are the same in each framework.

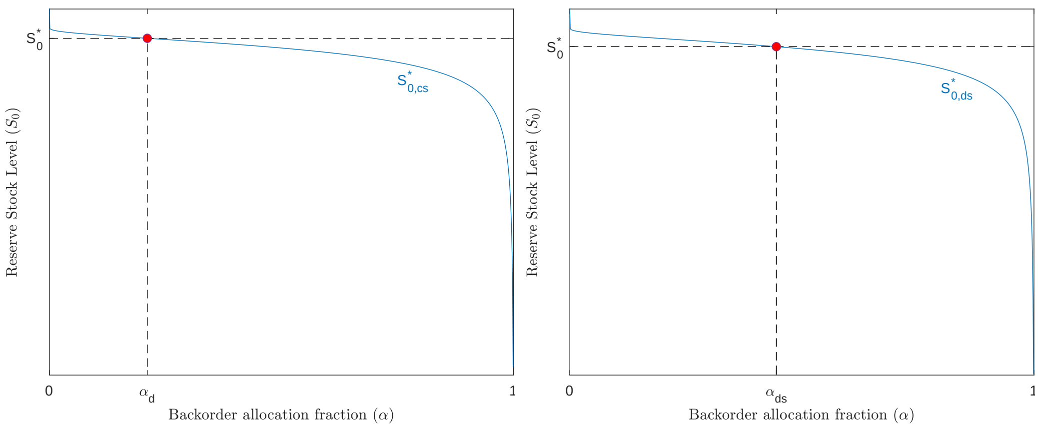
4.3.2 *Optimal Base-Stock Level*

In this case we compare the safety stock S_0^* in both the centralized and decentralized suppliers with respect to the centralized solution. We remind that in both cases, this quantity is of the form $S_{0,d}^* = z_\alpha \sigma(\mu_d^*)$ where z_α is an increasing function in α while we showed that σ is a decreasing function in α. Again we let μ_d^* be the optimal capacity level for either one of the decentralized cases. Another thing to consider is that z_α can take both positive and negative value. The derivative of S_0 with respect to α is given by

$$\frac{\partial z_\alpha}{\partial \alpha}\sigma(\mu_d^*) + z_\alpha \frac{\partial \sigma(\mu_d^*)}{\partial \alpha}.$$

We know that the derivative of $z_\alpha > 0$ since it's an increasing function and $\sigma(\mu_d^*)$ is positive by definition. The derivative of $\sigma(\mu_d^*)$ is negative as we showed earlier that $\sigma(\mu_d^*)$ is a decreasing function. Then, for the derivative to be positive we need $z_\alpha < 0$. This is true only when $\alpha\, b < h$. When $\alpha\, b > h$, it is unclear whether the first term is larger then the second or not. For that we plot $S_{0,d}^*$ vs α and see it's trend. We also compare it with S_0^* and check the α^* at which they intersect.

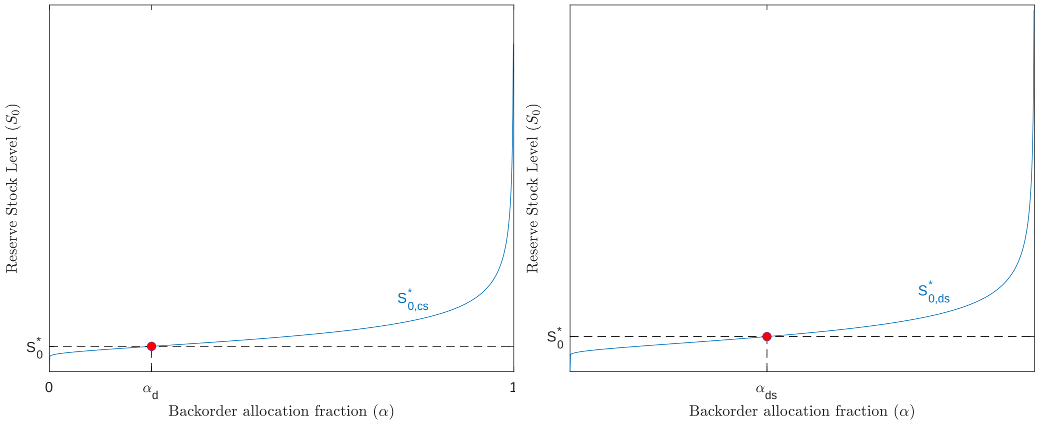
Figure 4.5: The Optimal Centralized and Decentralized Retailer's Base-Stock level $S_{0,cs}^*$ and $S_{0,ds}^*$ when $h > \alpha b$ as a Function of the Retailer's Backorder Share α —The Centralized Solution is S_0^*.



This verifies that when $h > \alpha b$, S_0^* is decreasing in both cases. These plots show the intersection between both decentralized cases and the centralized solution. Specifically, when $\alpha b < h$: $S_0^* < S_{0,d}^*$ for $\alpha < \alpha^*$ and $S_0^* > S_{0,d}^*$ for $\alpha > \alpha^*$. As for the values of α^*, we get that $\alpha_d = 0.2111$ and $\alpha_{ds} = 0.4450$. These values are obtained for $b = 10$ and $h = 14$ with $c = 3$. We change the values of b and h so that the condition $\alpha\, b < h$ is valid. Despite this change, the values of the α^*s are very close to the respective ones in the capacity and σ plots.

As for the second case, the plots are as follows.

Figure 4.6: The Optimal Centralized and Decentralized Retailer's Base-Stock level $S^*_{0,cs}$ and $S^*_{0,ds}$ when $\alpha b > h$ as a Function of the Retailer's Backorder Share α —The Centralized Solution is S^*_0.



It turns out that the derivative of S^*_0 is positive and this means that

$$\frac{\partial z_\alpha}{\partial \alpha}\sigma(\mu^*_d) > z_\alpha \frac{\partial \sigma(\mu^*_d)}{\partial \alpha}.$$

Moreover, When $\alpha b > h$: $S^*_0 > S^*_{0,d}$ for $\alpha < \alpha^*$ and $S^*_0 < S^*_{0,d}$ for $\alpha > \alpha^*$. As for the α^*, for the same parameters used for the capacity, we get $\alpha_d = 0.22$ and $\alpha_{ds} = 0.4242$ which, again, are equal to the previous ones.

4.3.3 *Optimal Total System Limiting Costs*

Finally, we look at the total costs. Specifically, we need to show that TC is strictly smaller than the decentralized costs. We start with the centralized suppliers' case. We need to show

$$\frac{3}{2}\Psi(z_0)^{2/3}c^{1/3} < \Psi^{2/3}(z_\alpha)c^{1/3}\left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3} + \frac{1}{2}\left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right).$$

It is clear that this is equivalent to that

$$\frac{3}{2}\left(\frac{\Psi(z_0)}{\Psi(z_\alpha)}\right)^{2/3} < \frac{2\Psi(z_\alpha) + \chi_\alpha}{2\chi_\alpha^{1/3}\Psi(z_\alpha)^{2/3}} \iff 3\Psi^{2/3}(z_0)\chi_\alpha^{1/3} < 2\Psi(z_\alpha) + \chi_\alpha.$$

In order to do so, we use the Geometric Mean Inequality (GMI). We remind that the GMI implies that for $a_i \geq 0$, $\frac{1}{n}\sum_{i=1}^{n} a_i \geq \sqrt{a_1\, a_2\, ..., a_n}$. For $n = 3$, let $a_1 = a_2 = Psi(z_0)$ and $a_3 = \chi_\alpha$. Then by the GMI we have

$$\frac{1}{3}(2\Psi(z_0) + \chi_\alpha) \geq (\Psi^2(z_0)\chi_\alpha)^{1/3}.$$

By cross multiplying we get

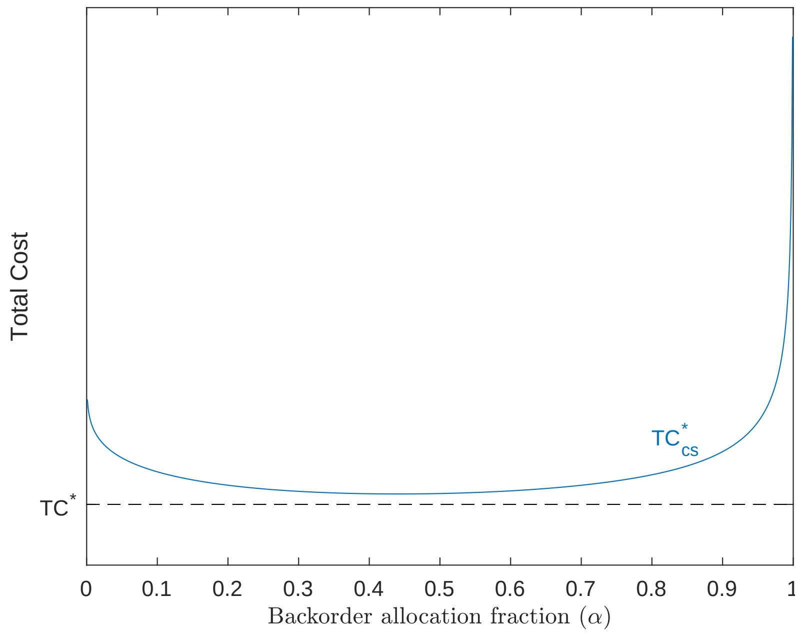
$$3\Psi^{2/3}(z_0)\chi_\alpha^{1/3} \leq 2\Psi(z_0) + \chi_\alpha.$$

One important element of the GMI is that the equality holds only when $a_i = a_j$ for $i \neq j$. Hence,

$$3\Psi^{2/3}(z_0)\chi_\alpha^{1/3} < 2\Psi(z_0) + \chi_\alpha.$$

Then we know that the total centralized limiting cost is smaller than the centralized suppliers' one.

Figure 4.7: The Optimal Centralized Suppliers' Total Cost $\widetilde{\mathrm{TC}}_{\mathrm{d}}^{*}$ as a Function of the Retailer's Backorder Share α —The Optimal Centralized System Cost is $\widetilde{\mathrm{TC}}^{*}$.



The plot above validates our result. In addition to showing that the centralized cost is always smaller than the centralized suppliers, it also confirm that they never intersect, hence showing that the centralized solution is the most efficient (as expected). As for the decentralized suppliers, we need to show that

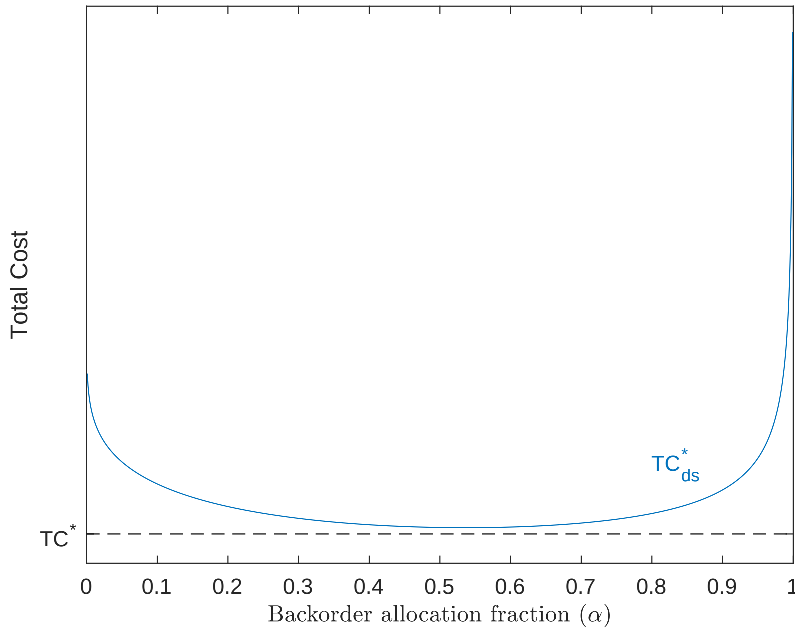
$$\frac{3}{2}\Psi(z_0)^{2/3}c^{1/3} < \frac{c^{1/3}\Psi(z_\alpha)^{2/3}}{2^{1/3}}\left(\left(\frac{\Psi(z_\alpha)}{\chi_\alpha}\right)^{1/3} + \left(\frac{\chi_\alpha}{\Psi(z_\alpha)}\right)^{2/3}\right).$$

This is equivalent to

$$3\Psi(z_0)^{2/3}\chi_\alpha^{1/3} < 2^{2/3}(\Psi(z_\alpha) + \chi_\alpha).$$

It does seem that this inequality holds as it is close to the one we proved previously, but as we can't assess whether $2^{2/3}(\Psi(z_\alpha) + \chi_\alpha) < 2\Psi(z_\alpha) + \chi_\alpha$ and we cant apply GMI here, instead we shift our analysis to a numerical one and plot both costs with respect to α.

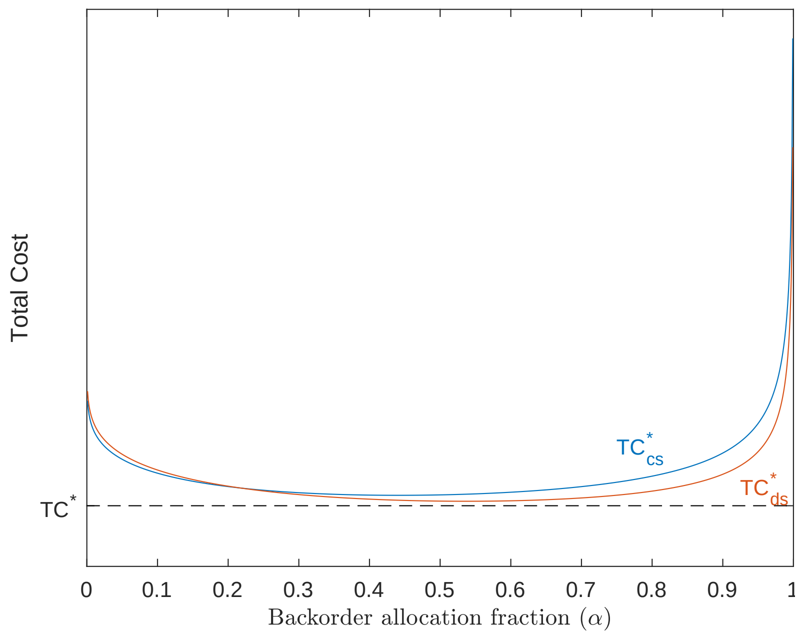
Figure 4.8: The Optimal Decentralized Suppliers' Total Cost $\widetilde{\mathrm{TC}}_{ds}^*$ as a Function of the Retailer's Backorder Share α —The Optimal Centralized System Cost is $\widetilde{\mathrm{TC}}^*$.



In this case, the decentralized limiting cost is closer to the the centralized cost than the centralized suppliers' one, however it is still clear that the strict inequality holds. This shows that the centralized solution is better than both decentralized ones.

During our analysis of above we established that it is not clear which of the decentralized costs is smaller. We look at both curves with respect to α and compare them. We keep our benchmark for reference.

Figure 4.9: The Optimal Centralized and Decentralized Limiting Costs as a Function of the Retailer's Backorder Share α —The Centralized Solution is $\widetilde{\text{TC}}^*$.



We can see that for $\alpha < \alpha^*$ the centralized suppliers' solution give a lower cost than the decentralized suppliers' solution. However, the costs are extremely close to each other. This $\alpha^* = 0.2170$ for the same parameters used in this analysis. On the other hand, for $\alpha > 0.2170$, there is higher difference between the two costs. It is then more beneficial for the suppliers to act independently, as their decentralized cost is smaller than the centralized one and even when it is higher there isn't much difference.

Chapter 5

Contracts

A supply chain is a multi-agent problem. As long as the information is symmetric among the various agents, a decentralized system is at best as efficient as the centralized one, the same way a local optimization is at best as effective as a global optimization. The distributed supply chain we are analyzing in this work is no exception. We have showed in Section 4.3 that the decentralized system is strictly inefficient relatively to the centralized one. This was also the case for the fully decentralized case compared to a decentralized supply chain but with a centralized supply function. These inefficiencies are also lost opportunities, possibly for all the players, to improve their utility function (here cost functions). To remedy this, decision makers have to carefully design mechanisms or contracts between these various players with the objective to reduce partially or fully these inefficiencies by eventually creating the right incentives for the players to select an optimal global solution known as the first-best solution.

We introduce in this section the notion of coordinating contract that specifies linear transfer payments between the supply chain players. We design these in a way that even though the retailer and suppliers would be individually selecting what is in their best interest, they end up due generating together the globally optimal cost. Contracts are a fundamental area in supply chain management literature as they govern the interactions between the various entities in the chain. Assuming simple cost functions that govern the interaction between these players is known to generate inefficiencies (e.g, the double marginalization effect). The literature is extremely rich and broad around this topic. We refer the reader to the review of [22] who also shares an overview of different contract types in the multi-echelon inventory context.

5.1 Contract structure

We focus specifically on the design and implementation of contracts with linear transfer payments in the context of distributed system. Linear transfer payments refer to direct monetary transfers between the retailer and suppliers based typically on predetermined criteria, such as order quantities, delivery schedules, or product quality. Unlike more complex incentive structures, linear transfer payments offer simplicity and transparency, making them particularly attractive for managing relationships in supply chains with multiple suppliers.

The transfer payments depend on both the order-up-to policy and the capacity levels selected respectively by the retailer and the suppliers. The cost functions of the retailer and the supply function are modified by this transfer payment $\tau(z, \boldsymbol{\mu})$ as follows:

$$\widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) = \widetilde{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) + \tau(z, \boldsymbol{\mu})$$

and

$$\widehat{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) = \widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) - \tau(z, \boldsymbol{\mu})$$

For such mechanism to generate the centralized solution, it must be that the final decisions of z and $\boldsymbol{\mu}$ in this context are such as:

$$\widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) + \widehat{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) = \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*).$$

Now, one way to achieve this (if at all possible) is by splitting $\widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*)$ between the retailer's and the supply function's costs. We denote by γ the splitting factor so that $\widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) = \gamma\, \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*)$ and $\widehat{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) = (1-\gamma)\, \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*)$. By doing so, we can easily show that that the transfer payment τ needs to be equal to

$$\tau(z, \boldsymbol{\mu}) = \gamma \widetilde{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) - (1-\gamma) \widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z). \tag{5.1}$$

Note that in this transfer of payment, the retailer is making a payment to the supply function and the supply function is making another payment to the retailer and τ is the net value paid (or received depending on the sign) by the retailer.

Obviously, by injecting this specific transfer payment in the expected total cost rate of the retailer and respectively, the supplier, we get that:

$$\widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) = \widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) - \gamma \widehat{\mathrm{TC}}_{\mathrm{r}}(z; \boldsymbol{\mu}) + (1-\gamma) \widehat{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) = (1-\gamma)\, \widetilde{\mathrm{TC}}(z, \boldsymbol{\mu}).$$

Similarly,

$$\widehat{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}; z) = \gamma\, \widetilde{\mathrm{TC}}(z, \boldsymbol{\mu}).$$

The existence of a unique Nash equilibrium as discussed previously, guarantees that such transfer payments insures that the retailer and the suppliers select the optimal global solution. However, for players to accept such mechanism design, they need to have the incentive to do so, i.e. are better off with this mechanism than without it. Basically:

$$\widetilde{\mathrm{TC}}_{\mathrm{r}}(z_d^*; \boldsymbol{\mu}_d^*) \geq (1-\gamma) \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*),$$

and

$$\widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}_d^*; z_d^*) \geq \gamma \widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*),$$

where $\boldsymbol{\mu}_d^*$ is the optimal decentralized capacity in either the centralized or decentralized suppliers' case. The problem is then reduced to showing the existence of such factor $\gamma \in (0, 1)$.

From the previous chapter, we already know the values of all the quantities in the previous inequalities except for γ. The previous inequalities are therefore equivalent to having

$$\underline{\gamma} \leq \gamma \leq \bar{\gamma},$$

where

$$\underline{\gamma} = 1 - \frac{\widetilde{\mathrm{TC}}_{\mathrm{r}}(z_\alpha; \boldsymbol{\mu}_d^*)}{\widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*)} \text{ and } \bar{\gamma} = \frac{\widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}_d^*; z_\alpha)}{\widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*)}.$$

In conclusion, the existence of a $\tau(z, \mu)$ that insures full coordination of the supply chain through a linear transfer payment, is equivalent to finding $\gamma \in [\underline{\gamma}, \bar{\gamma}] \cap [0, 1]$.

5.2 Decentralized suppliers

We remind that for c constant, the optimal limiting cost of the decentralized suppliers

- $\widetilde{\mathrm{TC}}_{\mathrm{r}}(z_\alpha; \boldsymbol{\mu}_{ds}^*) = \Psi(z_\alpha) \left(\frac{c}{\chi_\alpha}\right)^{1/3}$,
- $\widetilde{\mathrm{TC}}_{\mathrm{s}}(\boldsymbol{\mu}_{ds}^*; z_\alpha) = 2^{2/3} \chi_\alpha^{2/3} \, c^{1/3}$.

And the optimal centralized limiting cost is given by

- $\widetilde{\mathrm{TC}}(z_0, \boldsymbol{\mu}^*) = \frac{3}{2} \Psi(z_0)^{2/3} \, c^{1/3}$.

We remind that our goal is to show the existent (or lack thereof) of γ. Given that γ is such that $\gamma \in [\underline{\gamma}, \bar{\gamma}] \cap [0, 1]$. For that we look at the boundaries, specifically we look at the limit of these boundaries. Before we proceed, we introduce a Lemma that will be used throughout our analysis

Lemma 4

$$\varphi(x) \sim x \, \bar{\Phi}(x) \text{ when } x \to +\infty$$

and

$$\varphi(x) \sim -x \Phi(x) \text{ when } x \to -\infty.$$

to simplify our work, we also use the Lemma 3 and he optimal solutions to write

$$\begin{aligned}
\mathbb{E}[Z - z_0]^+ &= \left(-z_0 + \frac{\varphi(z_0)}{1 - \Phi(z_0)}\right) (1 - \Phi(z_0)) \\
&= \left(-z_0 + \frac{\varphi(z_0)}{1 - \frac{b}{b+h}}\right) (1 - \frac{b}{b+h}) \\
&= -z_0 \left(\frac{h}{h+b}\right) + \varphi(z_0),
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{E}[Z - z_\alpha]^+ &= \left(-z_\alpha + \frac{\varphi(z_\alpha)}{1 - \Phi(z_\alpha)}\right)(1 - \Phi(z_\alpha)) \\
&= \left(-z_\alpha + \frac{\varphi(z_\alpha)}{1 - \frac{\alpha b}{\alpha b + h}}\right)(1 - \frac{\alpha b}{\alpha b + h}) \\
&= -z_\alpha \left(\frac{h}{h + \alpha b}\right) + \varphi(z_\alpha).
\end{aligned}
$$

For simplicity, let $\Psi_\alpha = \Psi(z_\alpha)$ and $\Psi = \Psi(z_0)$. Given the optimal limiting cost, $\underline{\gamma}$ and $\bar{\gamma}$ are of the form

$$
\underline{\gamma} = 1 - \frac{2^{2/3}}{3} \frac{\Psi_\alpha}{\chi_\alpha^{1/3} \Psi^{2/3}} \text{ and } \bar{\gamma} = \frac{2^{5/3}}{3} \left(\frac{\chi_\alpha}{\Psi}\right)^{2/3}.
$$

Note that both $\underline{\gamma}$ and $\bar{\gamma}$ are dependent on α, h and b. We look into the limits with respect to α and h (the limits for b and h are equivalent so we only look at the limits as $h \to 0$).

Starting off our analysis, we look at the limits of the components of $\underline{\gamma}$ and $\bar{\gamma}$, mainly χ_α, Ψ_α and Ψ. The last element is independent of α, so we will study its limits when we move on to $h \to 0$. We also re-write

$$
\begin{aligned}
\Psi_\alpha &= (\alpha b + h)\mathbb{E}[Z - z_\alpha]^+ + h z_\alpha = \alpha b \mathbb{E}[Z - z_\alpha]^+ + h\left(\max(Z - z_\alpha, 0) + z_\alpha\right) \\
&= \alpha b \mathbb{E}[Z - z_\alpha]^+ + h \mathbb{E} \max(Z, z_\alpha).
\end{aligned}
$$

Limit as $\alpha \to 0$

Starting with $\alpha \to 0$, we look at the limit if the basic elements. Since

$$
\frac{\alpha\, b}{\alpha\, b + h} \to 0,
$$

then $z_\alpha \to -\infty$ as it's the inverse cdf of a normal random variable. The, the limits of $\bar{\Phi}(z_\alpha)$ and $\varphi(z_\alpha)$ are 1 and 0 respectively. Additionally, since $z_\alpha \to -\infty$, then

$$
\mathbb{E} \max(Z, z_\alpha) \to \mathbb{E}\, Z \to 0.
$$

The last limit is because $Z \sim \mathcal{N}(0, \sigma^2)$.

Given these limits, we can now find the limit of Ψ_α. Using the form of $\mathbb{E}[Z - z_\alpha]^+$ we wrote above, we now have

$$
\Psi_\alpha = \alpha\, b \mathbb{E}[Z - z_\alpha]^+ = \alpha\, b\, (z_\alpha\, \bar{\Phi}(z_\alpha) + \varphi(z_\alpha)).
$$

We reduce our problem by letting $\alpha \to \frac{h}{b}\Phi(z_\alpha)$. That is because

$$
\frac{\alpha\, b}{\alpha\, b + h} = \Phi(z_\alpha) \iff \alpha = \frac{h}{b} \frac{\Phi(z_\alpha)}{\bar{\Phi}(z_\alpha)} \to \frac{h}{b} \Phi(z_\alpha).
$$

The limit is due to the fact that $\bar{\Phi}(z_\alpha) \to 1$.

Then,

$$\Psi_\alpha \to \frac{h}{b}\,\Phi(z_\alpha)\,b(-z_\alpha\bar{\Phi}(z_\alpha) + \varphi(z_\alpha)) \to -z_\alpha\,h\,\Phi(z_\alpha).$$

We now use Lemma 4 for $z_\alpha \to -\infty$ to write

$$-z_\alpha\,\Phi(z_\alpha) \to \varphi(z_\alpha).$$

Hence,

$$\Psi_\alpha \to h\,\varphi(z_\alpha).$$

Then, we conclude that for $\alpha \to 0$,

$$\Psi_\alpha \to 0.$$

We follow a similar procedure for χ_α. Given that

$$\chi_\alpha = (1-\alpha)\,b\,\mathbb{E}[Z - z_\alpha]^+,$$

we first note that it is obvious that $(1-\alpha)\,b \to b$ when $\alpha \to 0$. We know that $z_\alpha \to -\infty$, $\varphi(z_\alpha) \to 0$ and $\bar{\Phi}(z_\alpha) \to 1$. Then,

$$\mathbb{E}[Z - z_\alpha]^+ = -z_\alpha\,\bar{\Phi}(z_\alpha) + \varphi(z_\alpha) \to +\infty.$$

Hence, we conclude that as $\alpha \to 0$,

$$\chi_\alpha \to +\infty.$$

Now that we have all the components, we look at the limits of $\underline{\gamma}$ and $\bar{\gamma}$.
Since, $\Psi_\alpha \to 0$ and $\chi_\alpha \to +\infty$ then, $\frac{\Psi_\alpha}{\chi_\alpha^{1/3}} \to 0$. Finally, and with the fact Ψ is independent of α, we conclude that

$$\underline{\gamma} = 1 - \frac{2^{2/3}}{3}\frac{\Psi_\alpha}{\chi_\alpha^{1/3}\Psi^{2/3}} \to 1 \text{ and } \bar{\gamma} = \frac{2^{5/3}}{3}\left(\frac{\chi_\alpha}{\Psi}\right)^{2/3} \to +\infty.$$

Limit as $\alpha \to 1$

For $\alpha \to 1$, $\frac{\alpha\,b}{\alpha,b+h} \to \frac{b}{b+h}$. Then, $z_\alpha \to z_0$. This means that for such limit of α,

$$\Psi_\alpha \to \Psi.$$

Then,

$$\frac{\Psi_\alpha}{\Psi^{2/3}} \to \Psi^{1/3}.$$

Given that $z_\alpha \to z_0$, we now also know that

$$\mathbb{E}[Z - z_\alpha]^+ \to \mathbb{E}[Z - z_0]^+.$$

Since $\chi_\alpha = (1-\alpha)\, b\mathbb{E}[Z - Z_\alpha]^+$, then $\chi_\alpha \to 0$. Finally, we conclude that as $\alpha \to 1$,

$$\underline{\gamma} = 1 - \frac{2^{2/3}}{3} \frac{\Psi_\alpha}{\chi_\alpha^{1/3}\Psi^{2/3}} \to -\infty \text{ and } \bar{\gamma} = \frac{2^{5/3}}{3}\left(\frac{\chi_\alpha}{\Psi}\right)^{2/3} \to 0.$$

Limit as $h \to 0$
As for the limit as $h \to 0$, $\frac{\alpha\, b}{\alpha\, b + h} \to 1$. Then, $z_\alpha \to +\infty$. For such limit of z_α, we get $\varphi(z_\alpha) \to 0$ and $\bar{\Phi}(z_\alpha) \to 0$. Given that $\chi_\alpha = (1-\alpha)b\mathbb{E}[Z - z_\alpha]^+$, $\Psi_\alpha = (b+h)\mathbb{E}[Z - z_\alpha]^+ + z_\alpha\varphi(z_\alpha)$ and $\Psi = \Psi_{\alpha=1}$ with $\mathbb{E}[Z - z_\alpha]^+ = -z_\alpha\bar{\Phi}(z_\alpha) + \varphi(z_\alpha)$, we know that we can not find a limit for $\underline{\gamma}$ and $\bar{\gamma}$ through a similar procedure used for the limits of α. Then, we do the following;

Define $H(z) = \mathbb{E}[Z - z]^+ = -z\,\bar{\Phi}(z) + \varphi(z)$, with $H'(z) = -\bar{\Phi}(z)$. We are interested in the ratio:

$$r(h) = \frac{H(z_\alpha)}{H(z_0)} \text{ as } h \to 0.$$

We recall the following

- $z_\alpha = \Phi^{-1}(\frac{\alpha\, b}{\alpha\, b + h}) = \Phi^{-1}\left(1 - h/(\alpha\, b) + o(h)\right)$
- $\Phi(z_\alpha) = 1 - h/(\alpha b) + o(h)$
- $z'_\alpha = \frac{-\alpha\, b}{(\alpha\, b + h)^2}\, \frac{1}{\varphi(z_\alpha)}$ where z'_α is the derivative with respect to h.
- $H'(z_\alpha) = -z'_\alpha\, \bar{\Phi}(z_\alpha)$
- $\frac{\partial}{\partial h}\varphi(z_\alpha) = -z'_\alpha\, z_\alpha\, \varphi(z_\alpha)$

We that $H(z) \sim \varphi(z)/z^2$ as $z \to \infty$ by using l'Hopital's rule.

Proof: From Lemma 4, we know that $\varphi(z) = z\bar{\Phi}(z) + o(z\bar{\Phi}(z))$. Therefore,

$$H(z) = \varphi(z) - z\bar{\Phi}(z) = o(z\bar{\Phi}(z)).$$

We move now to show that $H(z)z^2/\varphi(z) \to 1$ as $z \to +\infty$.

Using l'Hospital rule, $H(z)z^2/\varphi(z)$ has the same limit as:

$$\frac{2z\, H(z) - z^2\bar{\Phi}(z)}{-z\varphi(z)} = \frac{2H(z) - z\bar{\Phi}(z)}{-\varphi(z)} = \frac{o(z\bar{\Phi}(z)) + z\bar{\Phi}(z)}{\varphi(z)} \to 1,$$

as $z \to \infty$. The last limit is again obtained using Lemma 4. ■

From l'Hospital's rule we know that $r(h)$ has the same limit as

$$\begin{aligned}\frac{z'_\alpha\,\bar\Phi(z_\alpha)}{z'_1\,\bar\Phi(z_1)} &= \frac{z'_\alpha\,h/(\alpha b)}{z'_1\,h/b}\\ &= \frac{1}{\alpha}\frac{z'_\alpha}{z'_1}\\ &= \frac{1}{\alpha}\frac{\alpha b\,\varphi(z_1)}{(\alpha b+h)^2}\frac{(b+h)^2}{b\,\varphi(z_\alpha)} = \frac{(b+h)^2}{(\alpha b+h)^2}\frac{\varphi(z_1)}{\varphi(z_\alpha)}.\end{aligned}$$

Applying again l'Hospital rule, we conclude that $r(h)$ has the same limit as $h\to 0$ as

$$\frac{1}{\alpha^2}\frac{-z'_1\,z_1\,\varphi(z_1)}{-z'_\alpha\,z_\alpha\,\varphi(z_\alpha)} = \frac{1}{\alpha^2}\frac{1/b}{1/\alpha b}\frac{z_1}{z_\alpha} = \frac{1}{\alpha}\frac{z_1}{z_\alpha}.$$

Denote by l the limit of any sub-sequence (in h) of z_1/z_α as $h\to 0$. If $l\in[0,1)$, then recall that z_α (and z_1) converge to infinity as $h\to 0$.

Observe that

$$\begin{aligned}\varphi(z_1)/\varphi(z_\alpha) &= \exp((z_\alpha^2 - z_1^2)/2)\\ &= \exp((z_\alpha - z_1)(z_\alpha+z_1)/2)\\ &\geq \exp((z_\alpha - z_1)z_\alpha)\\ &= \exp((1-z_1/z_\alpha)z_\alpha^2)\to\infty\end{aligned}$$

as $h\to 0$.

This contradicts that $r(h)\to l/\alpha$. A similar argument holds if $l>1$. We conclude that $z_1/z_\alpha\to 1$ as $h\to 0$ and hence $\varphi(z_1)/\varphi(z_\alpha)$ must converge to α as $h\to 0$. Recall that $\Psi_\alpha = (\alpha b+h)H(z_\alpha)+h\,z_\alpha$ and $\chi_\alpha = (1-\alpha)bH(z_\alpha)$. We conclude that

$$\begin{aligned}\left(\frac{\Psi_\alpha}{\chi_\alpha}\right)^{1/3}\left(\frac{\Psi_\alpha}{\Psi_1}\right)^{2/3} &= \left(\frac{(\alpha b+h)H(z_\alpha)+h\,z_\alpha}{(1-\alpha)bH(z_\alpha)}\right)^{1/3}\left(\frac{(\alpha b+h)H(z_\alpha)+h\,z_\alpha}{(b+h)H(z_1)+h\,z_1}\right)^{2/3}\\ &= \left(\frac{(\alpha b+h)H(z_\alpha)+\alpha b\varphi(z_\alpha)+o(\varphi(z_\alpha))}{(1-\alpha)bH(z_\alpha)}\right)^{1/3}\left(\frac{(\alpha b+h)H(z_\alpha)+\alpha b\varphi(z_\alpha)+o(\varphi(z_\alpha))}{(b+h)H(z_1)+\alpha b\varphi(z_\alpha)+o(\varphi(z_\alpha))}\right)^{2/3}.\end{aligned}$$

The last equality results from the facts that $\alpha b\bar\Phi(z_\alpha) = h+o(h)$ and Lemma 4. Recall that $H(z) = o(\varphi(z))$ as $z\to\infty$, from which we conclude that the first term in the product following $2/3$ is equivalent to

$$\left(\frac{\alpha b}{1-\alpha}+\frac{\alpha}{1-\alpha}\frac{\varphi(z_\alpha)}{H(z_\alpha)}\right)^{1/3}\to+\infty.$$

As for the second term, it is equivalent to

$$\left(\frac{\alpha H(z_\alpha)+\alpha\varphi(z_\alpha)}{H(z_1)+\varphi(z_1)}\right)^{2/3}\sim\left(\frac{\alpha\varphi(z_\alpha)}{\varphi(z_1)}\right)^{2/3}\to 1.$$

Then the limits of $\underline{\gamma}$ and $\bar{\gamma}$ as $h \to 0$ are

$$\underline{\gamma} \to -\infty \text{ and } \bar{\gamma} \to 0.$$

We summarize the limits in the following proposition

Proposition 7 *The limits of $\underline{\gamma}$ and $\bar{\gamma}$ are given by*

- *As $\alpha \to 0$,*
$$\underline{\gamma} \to 1 \text{ and } \bar{\gamma} \to +\infty$$
- *As $\alpha \to 1$,*
$$\underline{\gamma} \to -\infty \text{ and } \bar{\gamma} \to 0.$$
- *As $h \to 0$,*
$$\underline{\gamma} \to -\infty \text{ and } \bar{\gamma} \to 0.$$

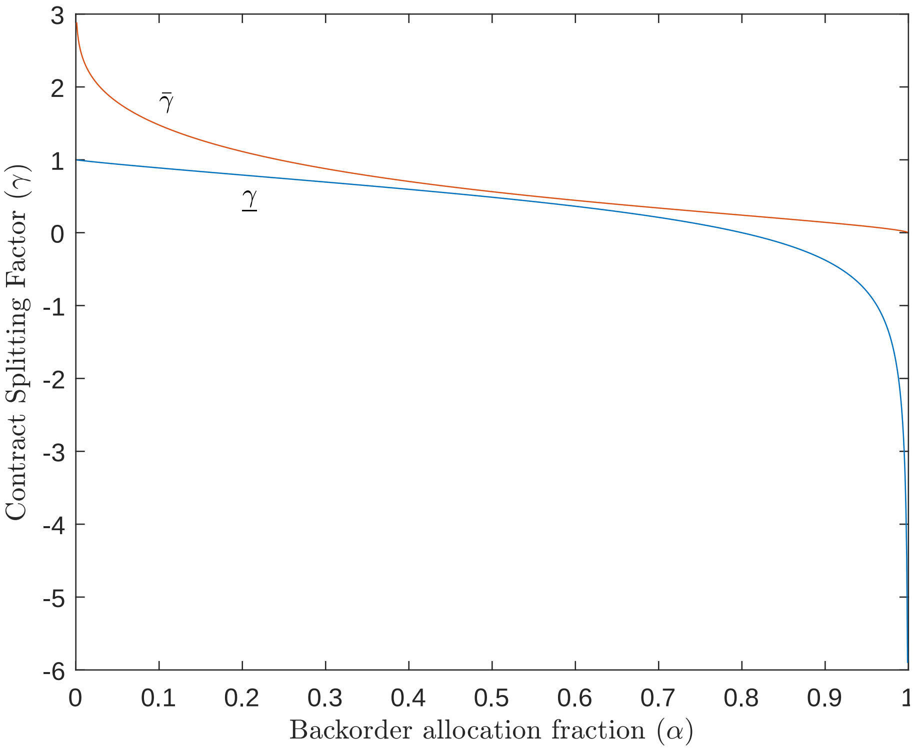
We mentioned previously that for use to have a valid contract, we need to show the existence of γ. Given the limits in the above proposition, and the fact that γ needs to be in $(0, 1)$, then we use the following proposition to show that we can always find a γ that satisfies the necessary conditions.

Proposition 8 *The existence of γ is guaranteed as $h \to 0$ and $\alpha \to \{0, 1\}$:*

- *For $0 < \alpha < 1$ small enough,*
$$0 < \underline{\gamma} < 1 < \bar{\gamma}.$$
- *For $0 < \alpha < 1$ large enough,*
$$0 < \underline{\gamma} < 0 < \bar{\gamma} < 1.$$
- *For $h > 0$ small enough,*
$$0 < \underline{\gamma} < 0 < \bar{\gamma} < 1.$$

Note that this also satisfies the fact that $\underline{\gamma}$ needs to be smaller then $\bar{\gamma}$. We verify our results by numerically plotting $\underline{\gamma}$ and $\bar{\gamma}$ with respect to α.

Figure 5.1: Decentralized Suppliers' γ's



5.3 Centralized Suppliers

Given the optimal limiting costs,

$$\underline{\gamma} = 1 - \frac{2}{3}\frac{\Psi_\alpha}{\chi_\alpha^{1/3}\Psi^{2/3}} \qquad \text{and} \qquad \bar{\gamma} = \left(\frac{\chi_\alpha}{\Psi}\right)^{2/3}.$$

It is clear that this boundaries are of the same form as the ones in the decentralized suppliers' case with a slight different being the coefficient. Since the limits of the ratios $\frac{\Psi_\alpha}{\chi_\alpha^{1/3}\Psi^{2/3}}$ and $\frac{\chi_\alpha}{\Psi}$ are go to either 0 or $+\infty$, then the limits will not be affected by these coefficients. Hence, all the results around showing the existence of γ still hold. We briefly summarise them in the following proposition.

Proposition 9 *The existence of γ is guaranteed as $h \to 0$ and $\alpha \to \{0, 1\}$:*

- *For $0 < \alpha < 1$ small enough,*

$$0 < \underline{\gamma} < 1 < \bar{\gamma}.$$

- *For $0 < \alpha < 1$ large enough,*

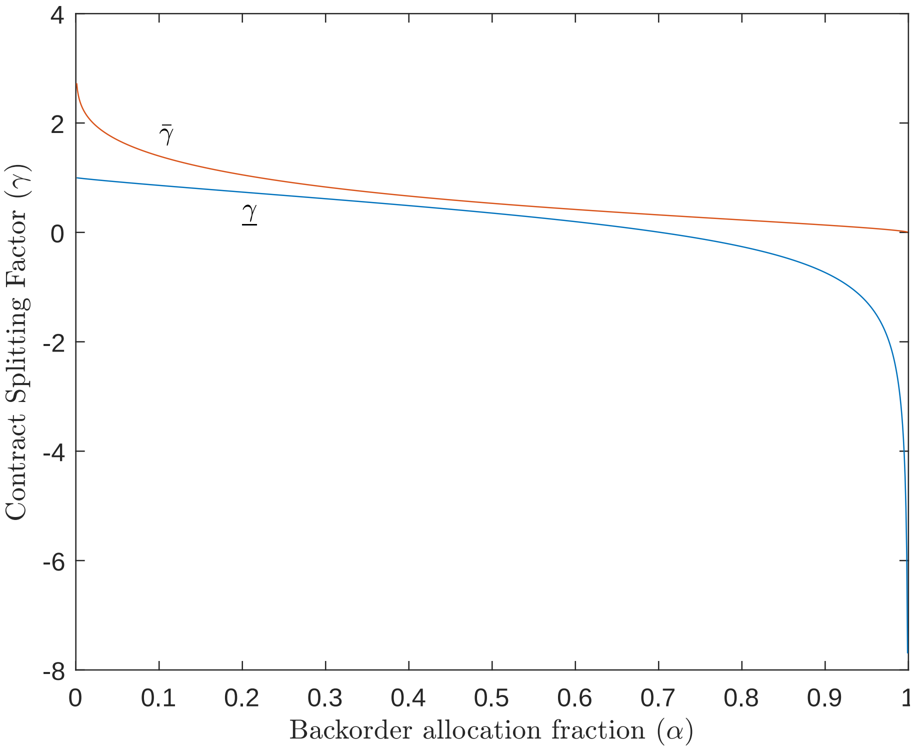
$$0 < \underline{\gamma} < 0 < \bar{\gamma} < 1.$$

- *For $h > 0$ small enough,*

$$0 < \underline{\gamma} < 0 < \bar{\gamma} < 1.$$

Again, we numerically verify then results with the plot below.

Figure 5.2: Centralized suppliers γ's



Chapter 6

The Stackelberg Games

In the previous chapters, our analysis revolved around the Nash equilibrium framework, where the agents simultaneously optimized their decision variables, thus minimizing their respective costs. Consequently, we now wish to change the system's dynamics by introducing a hierarchical structure with a designated leader (either the retailer or the supply function) in what is know as the Stackelberg games. Unlike the Nash equilibrium, where the agents act independently, the leader in the Stackelberg games has the advantage of making their decision first, with full knowledge of the followers' response. Followers would then adjust their decision accordingly. This creates an asymmetry in the decision-making process where the leader's strategy has an influence on the overall outcome of the game. This framework allows us to explore the leader's impact over the followers, potentially leading to different outcomes than situations where all agents share the same power. We then compare these outcomes and see if one framework offers an advantage.

6.1 Suppliers' Stackelberg Game

When the Suppliers are the leaders they choose their individual capacity levels $\mu_k^{*,n}$ to optimize $\mathrm{TC}_{\mathrm{s}}^{k,n}$ given the retailer's best response $S^{*,n}$ that minimized $\mathrm{TC}_{\mathrm{r}}^{n}$.

Proposition 10 *When $n \to +\infty$, and if $\xi_1 \sim \exp(\mu)$, for any supplier k,*

$$S_{0,sl}^{*,n} \to S_{0,sl}^{*} = \sigma(\mu) z_\alpha$$

where $z_\alpha = \Phi^{-1}\left(\frac{\alpha b}{\alpha b + h}\right)$, and

$$\mu_{k,sl}^{*} = \frac{\left(\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1}\right)^{1/3}}{\sqrt{\nu_k}\left(\mathbb{E}\sqrt{\hat{\nu}}\right)^{2/3}}$$

with $\hat{\nu} = \hat{c}/\chi_\alpha$ where χ_α is the optimal decentralized standardized cost, such that $\chi_\alpha = (1-\alpha) b \mathbb{E}[Z - z_\alpha]^+$.

Proof: Since this is the suppliers' game, then the supply function will optimize their capacity based on the decision of the retailer. Hence, we start by finding the optimal base stock level that minimizes the retailer's cost. Similarly to Chapter 4, we write the retailer's cost as follows:

$$\begin{aligned}\frac{\mathrm{TC}^n_{\mathrm{r}}(S^n;\boldsymbol{\mu}^n)}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\left(h\,\mathbb{E}[S^n_d - n\,\mathbb{E}\zeta - (N^n - n\mathbb{E}\zeta)]^+ + b\,\mathbb{E}[N^n - n\mathbb{E}\zeta - (S^n_d - n\mathbb{E}\zeta)]^+\right)\\ &= h\sigma\,\mathbb{E}[S^n_{0,d}/\sigma - X^n]^+ + b\,\sigma\,\mathbb{E}[X^n - S^n_{0,d}/\sigma]^+\end{aligned}$$

where we set $X^n = \frac{N^n - n\mathbb{E}[\zeta]}{\sigma\sqrt{n}}$. This is a newsvendor-like formulation with a constant term σ, and hence the optimal centralized stock, $S^{*,n}_{0,stack}/\sigma = (G^n)^{-1}(z_\alpha)$ where G^n is the cumulative distribution of X^n/σ. By Proposition 1 we have that $G^n(\cdot) \to \Phi(\cdot)$ which completes the first part of the proof. As for the capacity level, each suppliers we be minimizing the limiting cost equal to

$$\chi_\alpha \frac{\mu_k^{-1}}{\mathbb{E}_\Gamma \hat{\mu}^{-1}}\sigma(\boldsymbol{\mu}) + c_k\mu_k.$$

This has the exact same form as the limiting cost minimized by the decentralized suppliers. Hence, the proof follows form Chapter4 and concluding the second part of the proof. ■

We note that this is exactly equal to the decentralized suppliers' solution derived in Chapter 4.

6.2 Retailer's Stackelberg's Game

In this second case, we look at the game where the retailer chooses his $S^{*,n}$ to minimize $\mathrm{TC}^n_{\mathrm{r}}$ given the optimal μ^*_k chosen by each supplier. This case is much less tractable than the other game.

Proposition 11 *if* $\xi_1 \sim \exp(\mu)$, *for any supplier* k, *then,*

$$\mu^*_{k,rl} = \frac{\left(\mathbb{E}[\hat{\nu}^{1/2}] - \mathbb{E}[\hat{\nu}^{-1/2} + \hat{\nu}'^{-1/2}]^{-1}\right)^{1/3}}{\sqrt{\nu_k}\left(\mathbb{E}\sqrt{\hat{\nu}}\right)^{2/3}}$$

with $\hat{\nu} = \hat{c}/\chi(z)$ *where* $\chi(z)$ *is the optimal decentralized standardized cost, such that* $\chi(z) = (1-\alpha)b\mathbb{E}[Z-z]^+$.

As for the optimal base-stock level, the limiting cost of the retailer is now given by

$$\widetilde{TC}_{r,rl} = \frac{\psi(z)}{\chi(z)^{1/3}}\left((\mathbb{E}\sqrt{\hat{c}})^2 - \mathbb{E}\sqrt{\hat{c}}\,\mathbb{E}\frac{\sqrt{\hat{c}\hat{c}'}}{\sqrt{\hat{c}}+\sqrt{\hat{c}'}}\right)^{1/3} = \frac{\Psi(z)}{\chi(z)^{1/3}}C.$$

One would typically look at the first order condition, in a similar fashion as we solved the Suppliers' games. The F.O.C is given by

$$\frac{\partial\widetilde{TC}_{r,rl}}{\partial z} = \left(-\bar{\Phi}(z)(b+h) + h + \frac{1}{3}\frac{\Psi(z)}{\chi(z)}(1-\alpha)\,b\bar{\Phi}(z)\right)\frac{C}{\chi(z)}.$$

We need to find the vale of z that makes the F.O.C equal to 0. However, this problem is mathematically intractable and we cannot find a closed form for z. As such we shift our attention to a numerical approach. Our goal is to simulate the limiting cos of the retailer for a set of z and find the optimal one that minimized the cost. For that, we use the following approach:

- Generate a vector of z from $-\infty$ to ∞, with small increments
- Define the limiting cost as a function of z and find its respective value for the vector of z
- Find the minimum value of the cost and its respective z.

We note that we have 4 parameters, α, b, h and $\hat{c}$. As this is a decentralized framework, we fix b, h and $\hat{c}$ and find the optimal z for different values of α. We should first note that we take $\hat{c}, \hat{c}' \sim Unif(0,1)$, and find the respective value of C. For $b = 10$ and $h = 7$, the table below shows some of the values of z with respect to the parameters. We notice that both z_α and z_{stack} are increasing functions in α. We

α	0	0.1	0.5	0.7	0.99	1
z_{stack}	$-\infty$	-1.4056	-0.5701	-0.3839	-0.1910	$-\infty$
z_α	$-\infty$	-1.1503	-0.2104	0	0.2168	0.2230

Table 6.1: Optimal base-stock levels

also note that for $\alpha = 1$, z_{stack} jumps to $-\infty$ will z_α is positive.

We now compare the total limiting costs between the retailer and the suppliers' games. Note that by comparing these two limiting costs, we would also be comparing the Stackelberg solutions to the decentralized solution as we already established that when the suppliers' are the leaders the solution is equivalent to that of the decentralized suppliers.

We use the numerical solution we found to compute the limiting total cost of the supply chain and compare it to the total limiting decentralized suppliers' cost of the supply chain. We also compare them with our benchmark (centralized solution). Again, we compare them with respect to α.

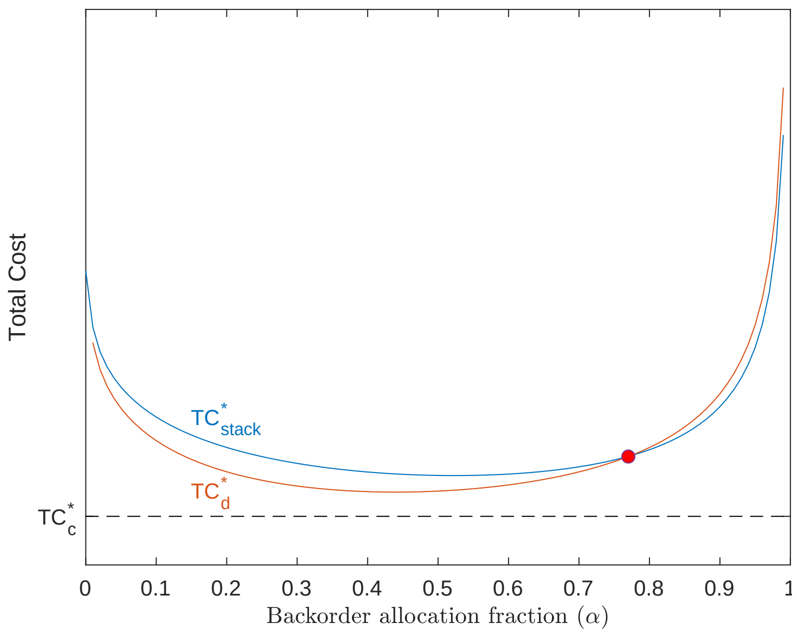


Figure 6.1: Stackelberg Total costs comparison

There are three things to note. First, the benchmark is the lowest cost for all values of α. This is expected, as the centralized solution is always the best. Second, we note that the total cost in the Retailer's game is higher then the one in the Suppliers' game (and decentralized suppliers $\mathrm{TC}_{\mathrm{d}}^{*}$) up to a certain value of α that we will call α_s. After this α_s, the total cost of the retailer's game will be lower than the other one, however the difference is smaller than the difference for$\alpha < \alpha_s$. Also note that as we approach $\alpha = 1$, the costs get closer to each other. Third, we use our numerical results to see the value of α_s. For the values of b, h and distribution of $\hat{c}$ chosen above, we have $\alpha_s = 0.77$.

Chapter 7

Numerical Analysis

In this section we discuss the numerical analysis we undertook. Throughout this thesis, we used both theoretical and numerical methods to solve our problems. We also used graphs to visually confirm any theoretical solution. However, in this chapter we will mainly rely on Monte Carlo simulation to validate our Central Limit Theorem-like result

7.1 Ns and Central Limit Theorem

We remind that what we need to validate is that in fact,

$$\frac{N_s^n - n\mathbb{E}[\xi_1]}{\sqrt{n}} \Rightarrow Z \sim \mathcal{N}(0, \sigma^2).$$

The first thing to do is to start by simulating N_s^n. Recall that

$$N_s^n \stackrel{d}{=} \sum_{j=0}^{\infty} I(n\xi_j + U > j).$$

One way to generate it as follows:

- Generate an arrival vector j that starts at 0 and grows very large, with $1/n$ increments.
- Generate a perturbation matrix ξ, with N rows and T rows, where N is the number of simulations and T is the length of the time vector j.
- Generate a vector U from a uniform distribution between 0 and 1. We remind that this is only used for technical reasons (to guarantee the existence of a steady state).

However, this method is quite heavy as generating a time vector that takes t to ∞ while making $n \to \infty$ will result in a very large j vector and even larger matrix ξ. To remediate this issue, we note that one can write N_s^n in a different way. Starting with the inside of the sum, one can write

$$I(n\xi_j > j - U) = \{1 \text{ for } j < \xi_j + U\}.$$

This is exactly equal to $\lfloor n\xi_j + U \rfloor$. Since

$$\frac{\lfloor n\xi_j + U \rfloor}{n} \overset{a.s.}{\to} \xi_j.$$

Then,

$$\frac{\mathbb{E}[\lfloor n\xi_1 + U \rfloor] + 1}{n} \to \mathbb{E}\xi_1. \tag{7.1}$$

We use the same steps described above while dropping the vector j and thus making ξ a vector and not a matrix. After generating the necessary variables, we proceed by calculating (7.1). We note that the expected value is over the number of simulations. For stability, one could make ξ an $N \times T$ matrix by where T does not grow large. We take $\mu = [1.2, 3.4]$, for $T = 50$, $N = 350$ and a vector $n = [1, 5, 10, 200, 1000]$, we show the convergence of $\frac{EN_s^n}{n}$ to $\mathbb{E}\xi_1 = \mu$ as n increases in the following tables.

$n = 1$	$n = 5$	$n = 10$	$n = 200$	$n = 1000$
2.2083	1.3881	1.3047	1.2194	1.2042

Table 7.1: Convergence for $\mu = 1.2$

$n = 1$	$n = 5$	$n = 10$	$n = 200$	$n = 1000$
4.3663	3.5651	3.4748	3.4112	3.4049

Table 7.2: Convergence for $\mu = 3.4$

$n = 1$	$n = 5$	$n = 10$	$n = 200$	$n = 1000$
8.2427	7.5250	7.3635	7.3214	7.3016

Table 7.3: Convergence for $\mu = 7.3$

Appendix A

Proofs

This section will discuss the proofs of some results.

Lemma 3.1. Proof:

$$\begin{aligned}
N_s^n &= \sum_{j=-\infty}^{t} I\left(j/n + \xi_j > t\right) \\
&\overset{d}{=} \sum_{j=-\infty}^{0} I\left(j/n + \xi_j > 0\right) \\
&= \sum_{j=0}^{\infty} I\left(-j/n + \xi_{-j} > 0\right) \\
&\overset{d}{=} \sum_{j=0}^{\infty} I\left(n\xi_j > j\right).
\end{aligned}$$

■

Proposition 1. Proof:

To show this, we use the moment generating function of Ns and compare it with the one for a Normal Random Variable. We start with

$$\begin{aligned}
\log \mathbb{E}\left[\frac{N_s^n - n\mathbb{E}\xi_1}{\sqrt{n}}\right] &= \log \mathbb{E}\left[\exp\left(\theta\frac{N_s^n}{\sqrt{n}}\right)\exp\left(\theta\frac{-n\mathbb{E}\xi_1}{\sqrt{n}}\right)\right] \\
&= \log\left[\exp(-\theta\sqrt{n}\mathbb{E}\,\xi_1)\mathbb{E}\left(\exp\theta\frac{N_s^n}{\sqrt{n}}\right)\right] \\
&= \log\left[\exp(-\theta\sqrt{n}\mathbb{E}\xi_1)\right] + \log\left[\mathbb{E}\exp\left(\theta\frac{N_s^n}{\sqrt{n}}\right)\right] \\
&= -\theta\sqrt{n}\mathbb{E}\xi_1 + \log\prod_{j=0}^{\infty}\mathbb{E}\exp\left(\frac{\theta}{\sqrt{n}}I_j^n\right)
\end{aligned}$$

Where $I_j^n = I\left(n\xi_j > j\right)$. We note that one can write

$$\mathbb{E}\left(\exp\left(\frac{\theta}{\sqrt{n}}I_j^n\right)\right) = \left(\exp(\frac{\theta}{\sqrt{n}} - 1)\mathbb{P}(n\xi_1 > j) + 1\right) = \left(\exp(\frac{\theta}{\sqrt{n}} - 1)\bar{F}(j/n) + 1\right).$$

This is done by noting that this is the moment generating function of an exponential random variable. Going back to our problem,

$$-\theta\sqrt{n}\mathbb{E}\xi_1 + \log\prod_{j=0}^{\infty}\mathbb{E}\exp\left(\frac{\theta}{\sqrt{n}}I_j^n\right) = -\theta\sqrt{n}\mathbb{E}\xi_1 + \log\prod_{j=0}^{\infty}\left(\exp(\frac{\theta}{\sqrt{n}} - 1)\bar{F}(j/n) + 1\right)$$
$$= -\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\log\left(\exp(\frac{\theta}{\sqrt{n}} - 1)\bar{F}(j/n) + 1\right)$$

Using Taylor Series approximation, we write

$$\exp(\frac{\theta}{\sqrt{n}}) - 1 \approx \frac{\theta^2}{2n} + \frac{\theta}{\sqrt{n}} + o(1/n).$$

Then,

$$-\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\log\left(\exp(\frac{\theta}{\sqrt{n}} - 1)\bar{F}(j/n) + 1\right)$$
$$= -\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\log\left(\frac{\theta^2}{2n}\bar{F}(j/n) + \frac{\theta}{\sqrt{n}}\bar{F}(j/n) + 1 + o(1/n)\right).$$

Again, using Taylor Approximation,

$$\log(x+1) \approx x - \frac{x^2}{2} + o(x^2).$$

We get,

$$-\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\log\left(\frac{\theta^2}{2n}\bar{F}(j/n) + \frac{\theta}{\sqrt{n}}\bar{F}(j/n) + 1 + o(1/n)\right)$$
$$= -\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\frac{\theta^2}{2n}\bar{F}(j/n) + \frac{\theta}{\sqrt{n}}\bar{F}(j/n) - \frac{1}{2}\left(\frac{\theta^2}{2n}\bar{F}(j/n) + \frac{\theta}{\sqrt{n}}\bar{F}(j/n)\right)^2 + o(1/n)$$
$$= -\theta\sqrt{n}\mathbb{E}\xi_1 + \sum_{j=0}^{\infty}\frac{\theta^2}{2n}\bar{F}(j/n) + \frac{\theta}{\sqrt{n}}\bar{F}(j/n)$$
$$-\frac{1}{2}\left(\frac{\theta^4}{4n^2}\bar{F}^2(j/n) + \frac{\theta^2}{n}\bar{F}^2(j/n) + \frac{\theta^3}{n\sqrt{n}}\bar{F}^2(j/n)\right) + o(1/n)$$

After some calculations, we can write the sum as

$$\sum_{j=0}^{\infty}\frac{\theta}{\sqrt{n}}\bar{F}(j/n) + \sum_{j=0}^{\infty}\frac{\theta^2}{2n}\left(\bar{F}(j/n) - \bar{F}^2(j/n) + o(1/n)\bar{F}^2(j/n)\right).$$

Note that $\sum_{j=0}^{\infty} \frac{\theta}{\sqrt{n}} \bar{F}(j/n) = \theta\sqrt{n}\mathbb{E}\xi_1$. This true through Riemann approximation, that changes the sum into integration;

$$\sum_{j=0}^{\infty} \bar{F}^2(j/n) \approx \int_1^{\infty} \bar{F}^2(x)dx = \mathbb{E}[\xi_1]^+.$$

Since $\xi > 0$, then this is equal to $\mathbb{E}\xi_1$. Then, our problem is reduced to

$$\sum_{j=0}^{\infty} \frac{\theta^2}{2n} \left(\bar{F}(j/n) - \bar{F}^2(j/n) + o(1/n)\bar{F}^2(j/n)\right) = \frac{\theta^2}{2n} \left(\bar{F}^2(j/n)F^2(j/n)\right) + o(1/n)\bar{F}^2(j/n)$$

Using Riemann approximation a second time, we get

$$\frac{\theta^2}{2n} \left(\bar{F}^2(j/n)F^2(j/n)\right) + o(1/n)\bar{F}^2(j/n) = \frac{\theta^2}{2} \int_0^{\infty} \bar{F}^2(x)F^2(x)dx.$$

Now that we have established the moment generating function of our problem, we note that for a normal random variable w e have

$$M_z(\theta) = \exp(\mu\theta + \sigma^2 \frac{\theta^2}{2}).$$

Hence,

$$\log M_z(\theta) = \mu\theta + \sigma^2 \frac{\theta^2}{2}.$$

By comparison, we note that they are equal for $\mu = 0$ and $\sigma^2 = \int_0^{\infty} \bar{F}^2(x)F(x)dx$.
■

Lemma 2 Proof:

Part 1.

By integration by part we have

$$\begin{aligned}
\sigma^2(\boldsymbol{\mu}) &= \mathbb{E}_\Gamma \int_0^{\infty} \int_0^{x} F(t)dt\, f(x|\hat{\mu})dx \\
&= \mathbb{E} \int_0^{\xi(\hat{\mu})} (1 - \mathbb{E}_\Gamma \mathbb{P}(\xi_1 > x|\hat{\mu}'))dx \\
&= \mathbb{E}\xi(\hat{\mu}) - \mathbb{E} \int_0^{\xi(\hat{\mu})} \mathbb{E}_\Gamma[e^{-\hat{\mu}'x}|\hat{\mu}']\, dx \\
&= \mathbb{E}[\hat{\mu}^{-1}] - \mathbb{E}\left[\hat{\mu}'^{-1}(1 - e^{-\hat{\mu}'\xi(\hat{\mu})})\right] \\
&= \mathbb{E}\left[\hat{\mu}'^{-1}\, e^{-\hat{\mu}'\xi((\hat{\mu})}\right] = \mathbb{E}_\Gamma[\hat{\mu}'^{-1}\, \mathbb{E}_{\xi(\hat{\mu})} e^{-\hat{\mu}'\xi}|\hat{\mu}] = \\
&= \mathbb{E}_\Gamma[\hat{\mu}'^{-1} \frac{\hat{\mu}}{\hat{\mu} + \hat{\mu}'}] = \mathbb{E}[\frac{1}{\hat{\mu}'}] - \mathbb{E}[\frac{1}{\hat{\mu} + \hat{\mu}'}].
\end{aligned}$$

Recall that $\mathbb{E}\xi = 1/\hat{\mu}$ (one realization of $\hat{\mu}$) and the minimum of two exponential r.v. has a rate equal to the sum of the rates of each exponential. Therefore, we can write

$$\mathbb{E}_\Gamma[\frac{1}{\hat{\mu}'}] - \mathbb{E}_\Gamma[\frac{1}{\hat{\mu} + \hat{\mu}'}] = \mathbb{E}_\Gamma[\mathbb{E}\xi - \mathbb{E}\min(\xi, \xi') = \mathbb{E}\max(0, \xi - \xi'),$$

where the latter expected value is on both following Γ and the distributions of the ξ's.

Part 2. We start by writing

$$\begin{aligned}\sigma^2 &= \mathbb{E}\left[\frac{1}{\hat{\mu}} - \frac{1}{\hat{\mu}+\hat{\mu}'}\right] = \mathbb{E}\left[\frac{\hat{\mu}'}{\hat{\mu}(\hat{\mu}+\hat{\mu}')}\right] \\ &= \mathbb{E}\left[\frac{1}{\hat{\mu}}\right]\mathbb{E}\left[\frac{\hat{\mu}'}{\hat{\mu}+\hat{\mu}'}\right] + \mathbb{C}\mathrm{ov}\left(\frac{1}{\hat{\mu}}, \frac{\hat{\mu}'}{\hat{\mu}+\hat{\mu}'}\right) \\ &\geq \frac{1}{2\,\bar{\mu}}.\end{aligned}$$

The last inequality is due to three interesting facts. First, Jensen's inequality shows that $\mathbb{E}\left[\frac{1}{\hat{\mu}}\right] \geq 1/\bar{\mu}$. Secondly, for any two r.v.'s X and Y that are i.i.d. we can show that

$$\mathbb{E}\left[\frac{X}{X+Y}\right] = \frac{1}{2}.$$

Indeed,

$$1 = \mathbb{E}\left[\frac{X+Y}{X+Y}\right] = \mathbb{E}\left[\frac{X}{X+Y}\right] + \mathbb{E}\left[\frac{Y}{X+Y}\right] = 2\,\mathbb{E}\left[\frac{X}{X+Y}\right].$$

The last equality is obtained by symmetry. Finally, we show that if f and g are both monotone decreasing or increasing then $f(X)$ and $g(X)$ are positively correlated for any r.v. X. Indeed, let Y and X i.i.d. We have that

$$(f(X) - f(Y))(g(X) - g(Y)) \geq 0.$$

Taking expected values, we have that

$$\begin{aligned}0 &\leq \mathbb{E}f(X)g(X) - \mathbb{E}f(X)g(Y) - \mathbb{E}f(Y)g(X) + \mathbb{E}f(Y)g(Y) \\ &= \mathbb{E}f(X)g(X) - \mathbb{E}f(X)\mathbb{E}g(X) - \mathbb{E}f(Y)\mathbb{E}g(Y) + \mathbb{E}f(Y)\,g(Y) \\ &= 2\,\mathbb{E}f(X)g(X) - 2\mathbb{E}f(X)\mathbb{E}g(X),\end{aligned}$$

which shows that $f(X)$ and $g(X)$ are positively correlated. Given that $\hat{\mu}$ and $\hat{\mu}'$ are independent, we condition on $\hat{\mu}'$ and conclude that $1/\hat{\mu}$ and $\frac{\hat{\mu}'}{\hat{\mu}+\hat{\mu}'}$ are positively correlated which shows that

$$\mathbb{C}\mathrm{ov}\left(\frac{1}{\hat{\mu}}, \frac{\hat{\mu}'}{\hat{\mu}+\hat{\mu}'}\right) \geq 0.$$

■

Lemma 3 Proof:

$$\begin{aligned}\mathbb{E}[Z - z_0]^+ = \mathbb{E}[X]^+ &= \mathbb{E}[X|X>0]\mathbb{P}(X>0) \\ &= \left(\mu_X + \sigma_X \frac{\varphi(\alpha)}{1-\Phi(\alpha)}\right)\mathbb{P}(Z > z_0) \\ &= \left(-z_0 + \frac{\varphi(z0)}{1-\Phi(z_0)}\right)(1 - \Phi(z_0))\end{aligned}$$

where $\alpha = \frac{a-\mu_X}{\sigma_X} = \frac{0-(-z_0)}{1} = z_0$ ■

Lemma 4 Proof: For $x \to \pm\infty$, $x\bar{\Phi}(x)/\varphi(x)$ has the same limit as

$$\frac{(x\bar{\Phi}(x))'}{\varphi(x)'} = \frac{\bar{\Phi}(x) - x\varphi(x)}{\varphi(x)'} = \frac{\bar{\Phi}(x) - x\varphi(x)}{x\varphi(x)} = \frac{\bar{\Phi}(x)}{-x\varphi(x)} + 1$$

Using l'Hopital's rule since this is a 0/0 limit, this equation has the same limit as

$$\frac{-\varphi(x)}{-\varphi(x) - x\varphi(x)'} = \frac{\varphi(x)}{\varphi(x) + x^2\varphi(x)} = \frac{1}{1+x^2} \to 1.$$

Hence, $\varphi(x) \sim x\,\bar{\Phi}(x)$ when $x \to +\infty$ and $\varphi(x) \sim -x\Phi(x)$ when $x \to -\infty$. ■

Bibliography

[1] D. Erlenkotter, "Ford whitman harris and the economic order quantity model," *Operations Research*, vol. 38, no. 6, pp. 937–946, 1990.

[2] A. J. Clark and H. Scarf, "Optimal policies for a multi-echelon inventory problem," *Management science*, vol. 6, no. 4, pp. 475–490, 1960.

[3] J.-S. Song, G.-J. van Houtum, and J. A. Van Mieghem, "Capacity and inventory management: Review, trends, and projections," *Manufacturing & Service Operations Management*, vol. 22, no. 1, pp. 36–46, 2020.

[4] J. A. Van Mieghem, "Commissioned paper: Capacity management, investment, and hedging: Review and recent developments," *Manufacturing & Service Operations Management*, vol. 5, no. 4, pp. 269–302, 2003.

[5] W. Jammernegg and G. Reiner, "Performance improvement of supply chain processes by coordinated inventory and capacity management," *International Journal of Production Economics*, vol. 108, no. 1-2, pp. 183–190, 2007.

[6] H. L. Lee, V. Padmanabhan, and S. Whang, "Information distortion in a supply chain: The bullwhip effect," *Management science*, vol. 43, no. 4, pp. 546–558, 1997.

[7] G. P. Cachon, "Supply chain coordination with contracts," *Handbooks in operations research and management science*, vol. 11, pp. 227–339, 2003.

[8] G. P. Cachon and P. H. Zipkin, "Competitive and cooperative inventory policies in a two-stage supply chain," *Management science*, vol. 45, no. 7, pp. 936–953, 1999.

[9] G. P. Cachon, "Stock wars: Inventory competition in a two-echelon supply chain with multiple retailers," *Operations Research*, vol. 49, no. 5, pp. 658–674, 2001.

[10] S. Axsäter, "A framework for decentralized multi-echelon inventory control," *IIE Transactions*, vol. 33, no. 2, pp. 91–97, 2001.

[11] G. P. Cachon and A. G. Kök, "Competing manufacturers in a retail supply chain: On contractual form and coordination," *Management science*, vol. 56, no. 3, pp. 571–589, 2010.

[12] P. Naor, "The regulation of queue size by levying tolls," *Econometrica: journal of the Econometric Society*, pp. 15–24, 1969.

[13] H. Mendelson and S. Whang, "Optimal incentive-compatible priority pricing for the m/m/1 queue," *Operations research*, vol. 38, no. 5, pp. 870–883, 1990.

[14] R. Caldentey and L. M. Wein, "Analysis of a decentralized production-inventory system," *Manufacturing & Service Operations Management*, vol. 5, no. 1, pp. 1–17, 2003.

[15] T. A. Taylor, "On-demand service platforms," *Manufacturing & Service Operations Management*, vol. 20, no. 4, pp. 704–720, 2018.

[16] V. F. Araman, A. Calmon, and K. Fridgeirsdottir, "Pricing and job allocation in online labor platforms," 2019.

[17] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[18] M. Huang, R. P. Malhamé, and P. E. Caines, "Large population stochastic dynamic games: Closed-loop mckean-vlasov systems and the nash certainty equivalence principle," *Communications in Information & Systems*, vol. 6, no. 3, pp. 221–252, 2006.

[19] M. Laurière, S. Perrin, M. Geist, and O. Pietquin, "Learning mean field games: A survey," *arXiv preprint arXiv:2205.12944*, 2022.

[20] S. R. Balseiro, O. Besbes, and G. Y. Weintraub, "Repeated auctions with budgets in ad exchanges: Approximations and design," *Management Science*, vol. 61, no. 4, pp. 864–884, 2015.

[21] V. F. Araman, H. Chen, P. W. Glynn, and L. Xia, "On a single server queue fed by scheduled traffic with pareto perturbations," *Queueing Systems*, vol. 100, no. 1-2, pp. 61–91, 2022.

[22] G. P. Cachon, "Competitive supply chain inventory management," in *Quantitative models for supply chain management*, Springer, 1999, pp. 111–146.